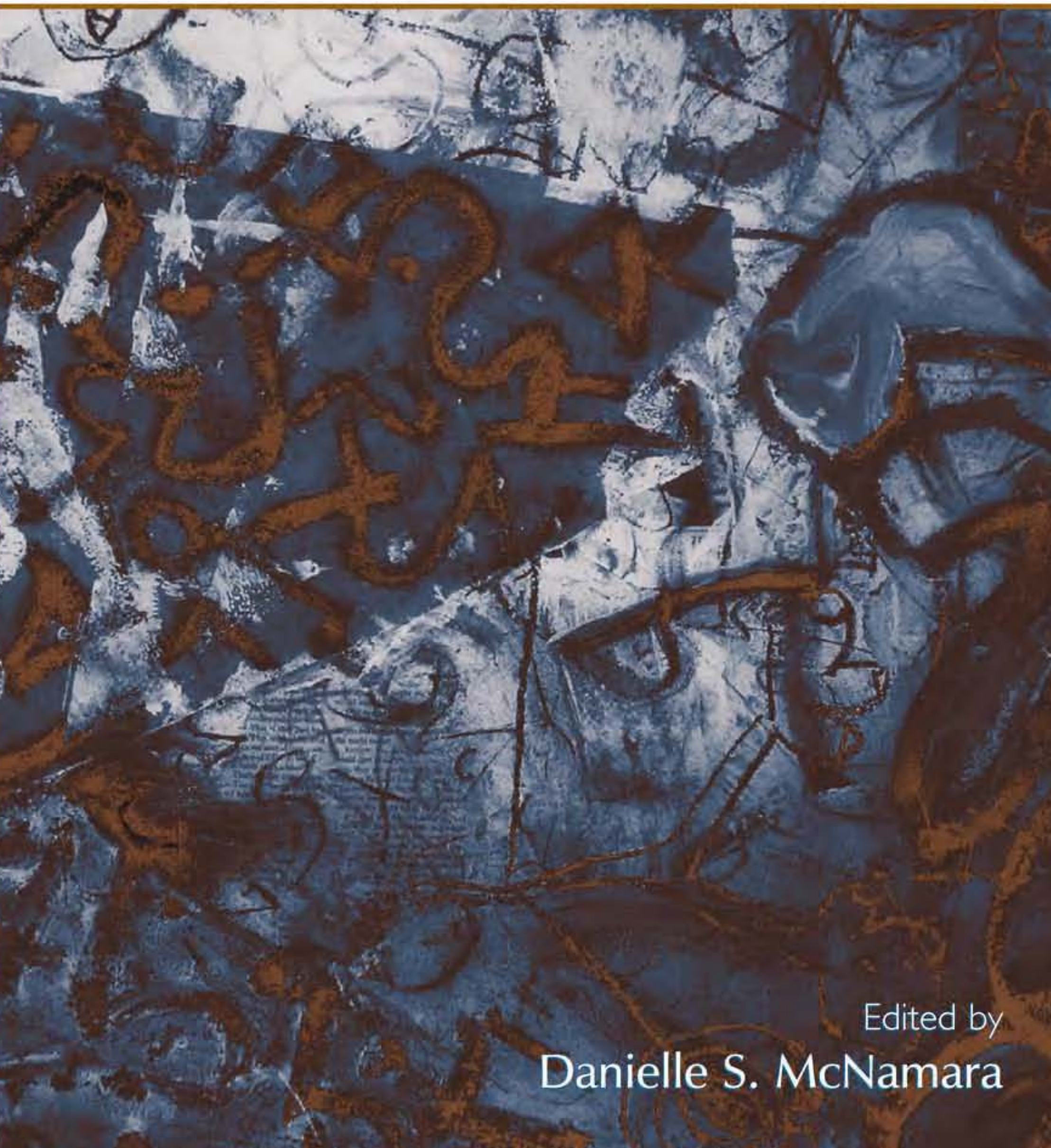
Reading Comprehension Strategies

Theories, Interventions, and Technologies



Edited by
Danielle S. McNamara

Reading Comprehension Strategies

Theories, Interventions, and Technologies

Reading Comprehension Strategies

Theories, Interventions, and Technologies

Edited by
Danielle S. McNamara
University of Memphis

Lawrence Erlbaum Associates
Taylor & Francis Group
New York London

Cover graphics by Randy McNamara:

Randy McNamara has an MFA from Indiana University, has won numerous awards, and has shown extensively on the east and west coasts, most recently at Gallery 825 in Los Angeles. He presently resides in Los Angeles with his wife and two young sons.

Front cover: "Moniker Mumble" (2006), 30" x 22"
Acrylics, pastel, and chalk on paper

Back cover: "Slogan Slur" (2006), 30" x 22"
Acrylics, pastel, and chalk on paper

Lawrence Erlbaum Associates, Inc., Publishers
10 Industrial Avenue
Mahwah, New Jersey 07430
www.erlbaum.com

Cover design by Tomai Maridou

Library of Congress Cataloging-in-Publication Data

Reading Comprehension Strategies: Theories, Interventions, and Technologies

ISBN 978-0-8058-5966-9 — ISBN 0-8058-5966-7 (cloth)
ISBN 978-0-8058-5967-6 — ISBN 0-8058-5967-5 (pbk)
ISBN 978-1-4106-1666-1 — ISBN 1-4106-1666-5 (e book)

Copyright information for this volume can be obtained by contacting the Library of Congress.

Books published by Lawrence Erlbaum Associates are printed on acid-free paper,
and their bindings are chosen for strength and durability.

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

My brother, the artist who created the cover of this book, and I, dedicate this book to the loving memory of our father, an avid reader who encouraged creativity of all kinds, both artistic and intellectual

Contents

Preface

What Is This About?

Reading can be challenging, particularly when the material is unfamiliar, technical, or complex. Moreover, for some readers, comprehension is *always* challenging. They may understand each word separately, but linking them together into meaningful ideas often doesn't happen as it should. These readers can *decode* the words, but have not developed sufficient skills to *comprehend* the underlying, deeper meaning of the sentences, the paragraphs, and the entire text. Comprehension refers to the ability to go beyond the words, to understand the ideas and the relationships between ideas conveyed in a text. The focus of this book is on the cognitive processes involved in comprehension, and moreover, on techniques that help readers improve their ability to comprehend text. The focus of this book is on *reading comprehension strategies.* Indeed, the use of effective reading comprehension strategies is perhaps the most important means to helping readers improve comprehension and learning from text.

There is a great deal of evidence for the importance of reading strategies. One source of evidence is that successful readers know when and how to use deliberate strategies to repair comprehension. One implication from that finding is that teaching reading strategies to struggling readers may be a key toward helping them to improve comprehension. And it is. Teaching struggling comprehenders to use strategies improves their comprehension and their ability to learn from challenging text. Thus, the use of reading strategies is an integral part of normal comprehension and teaching reading strategies should be an integral part of K–14 education.

What are reading comprehension strategies? To answer that question, let's start with cognitive learning strategies, such as mnemonics. Mnemonics help people to remember things such as lists of items, a speech, or lines in a play.

For example, one example of a mnemonic to aid memory of a list of items is to imagine a well known spatial route and visually place each item in a particular location along the route. Then, to recall the items, the person imagines traveling the route and *picking up* each item along the way. Another mnemonic, called *chaining,* is to create sentences out of the words in the lists. For example, with the words, *table, helicopter, saxophone,* and *leg* a sentence such as *'the table inside the helicopter had a saxophone for a leg'* would link the words visually, and thus the words would become more memorable. With practice, these types of memory aids can more than triple the number of items remembered. At first, these types of strategies take more time than just reading the list, but with practice, they become rapid, efficient, and effective—you remember more, with less effort. Likewise, reading strategies take more time at first, but with practice, help the reader to understand and remember much more from the text in less time than it would take without using reading strategies. For example, one reading strategy that pervades the literature is asking questions before, while, and after reading. At first, such a strategy will take the reader much more time and effort, and may even seem inefficient. But, with practice such strategies become more automatic, and then they become a natural part of reading. The focus of this volume is on why, when, and for whom such strategies are effective.

This volume provides an overview of reading comprehension strategies and strategy interventions that have been shown empirically to be effective in helping readers to overcome comprehension challenges. This volume differs from other books that might be found on reading strategies in two important ways. First, there is a heavy focus throughout on theories of reading comprehension: How well do current models of reading comprehension account for the importance of reading strategies? And most important, how do theories of reading comprehension motivate and support reading comprehension interventions?

Second, there is a focus on how current technologies can aid in helping teachers to provide reading strategy training to their students. One-on-one strategy training, and even focused group training is challenging for many teachers who are not specifically trained in reading and who don't have time to divert energy away from the teaching of critical content. New technologies are described that help the teacher be better prepared to engage their students in reading strategies in the classroom. And, computer-based tutoring technologies are described that offer further solutions to teachers' challenges by providing students with strategy training that can interact with and engage the student, and adapt to their individual needs.

What Is in This Volume?

This volume is divided into five sections. The first section includes four chapters that discuss theories of text comprehension, and in particular, the role that theories have played in identifying strategies that characterize expert reading and strategies that can be effectively taught. Art Graesser presents an overview of theories of reading comprehension, with an emphasis on the status of comprehension strategies within reading theories. Panayiota (Pani) Kendeou, Paul van den Broek and colleagues discuss the potential importance of pre-reading comprehension strategies. They argue convincingly that comprehension skills develop early in children's lives and that comprehension skills and basic reading skills (e.g., decoding) develop independently. The chapter by Jane Oakhill and Kate Cain carries forward that conclusion into early reading development. They present evidence that early competencies in skills related to inference making, comprehension monitoring, and understanding story structure causally influence comprehension development between the ages of 7 and 11, whereas skills related to decoding words have less influence on comprehension skill development. This section concludes with Michael Vitale and Nancy Romance's knowledge-based account of comprehension that argues for the embedding of reading strategy instruction within content area classes. They posit that promoting the use of reading strategies in meaningful, content specific learning environments is a more effective approach to enhancing reading comprehension proficiency than engaging students in a series of unrelated stories.

The second section looks at methods of using comprehension skill assessment to guide reading interventions. The chapter by Joe Magliano, Keith Millis and colleagues presents exciting new methods of automatically assessing deep level of comprehension by having students think aloud and answer questions while reading. They demonstrate that this type of method is more effective than more traditional standardized methods of assessment and shows greater promise in guiding individualized reading strategy interventions. The second chapter in this section, by Arthur Vander Veen, Kristen Huff and colleagues shows how a traditional, standardized method of measuring comprehension, the SAT, might nevertheless be used to guide comprehension interventions. Both of these chapters take novel approaches to comprehension assessment that are more tightly aligned with theories of reading comprehension and the critical role of reading strategies.

The third section delves into the heart of the matter, successful reading comprehension strategy interventions. Doug and Lynn Fuchs describe their intervention, called *Peer-Assisted Learning,* which entails pairing children from preschool through the intermediate elementary grades to engage in reading activities including repeated reading, paragraph summaries, and making predictions. Joanna Williams describes her text structure intervention that teaches second grade students how to use the structure of the text to better understand content area readings. Art Glenberg, Beth Jaworski and colleagues describe an intervention to enhance imagery processes for first- and second-grade children that involves either manipulating or imagining the process of manipulating toys that represent characters and objects in stories. John Guthrie, Ana Taboada and colleagues describe Concept-Oriented Reading Instruction (CORI), a broad strategy intervention for elementary school children that includes an emphasis on motivational practices for encouraging conceptual goal setting and affording student choice and collaboration. Finally, Alison King describes her intervention that helps elementary and middle school readers learn how to ask deep level questions while reading.

The fourth section of the book contains seven chapters on exciting new technologies that provide children with dynamic scaffolds toward active comprehension and help teachers learn how to provide strategy training. Mina Johnson-Glenberg describes her *3D-Readers* tutoring system that instructs and assesses elementary to middle school children in comprehension strategies such as visualization and question generation. Nicola Yuill's new software engages pairs of 7 to 9 year old children in discussing joking riddles that play on meanings of words, thus increasing children's awareness of inferences in text. Bonnie Meyer and Kay Wijekumar describe their tutoring system that teaches students to use knowledge about the structure of text while reading. Donna Caccamise, Marita Franzke and colleagues describe *Summary Street,* an interactive tutoring system that teaches middle school students how to summarize text more effectively through guided practice. Then, in chapter 16, I and my colleagues describe iSTART, a reading strategy tutor that teaches high school and college students how to self-explain text and use reading strategies such as making bridging inferences and elaborations while reading challenging text. Brigit Dalton and Patrick Proctor describe their use of universally designed digital literacy environments that scaffold reading strategy instruction for struggling elementary and middle school readers and students with learning disabilities. Finally, Annemarie Palincsar, Rand Spiro and colleagues describe their design of a hypermedia environment that uses new technologies to scaffold the use of videos to help teachers learn more effective techniques for providing children with reading comprehension instruction.

Section 5 is a concluding chapter by myself and my colleagues that presents the *4-Pronged Comprehension Strategy Framework.* This chapter organizes the various strategies described in this volume within a single framework and describes the theoretical and empirical rationale for the reading strategies included within the reading standards of the 2006 College Board English Language Arts College Board Standards for College Success™.

How Did This Volume Come About and Who Do We Have to Thank?

This volume was preceded in May 2005 by a workshop at the University of Memphis. We met there to discuss our research and to find common ground among reading theorists and researchers developing and testing reading strategy interventions. The workshop was immensely useful, illuminating, and fun. The workshop was partially funded by the Institute for Intelligent Systems and the Department of Psychology at the University of Memphis; I am extremely grateful for the University's support of research endeavors such as these. I am also grateful to many individuals who helped or organize that workshop, including the staff of the Institute of Intelligent Systems (Renee Cogar and Mattie Haynes) and the many student volunteers who helped make the conference a great success. I further thank the chapter reviewers for their dedication to the field. Although the chapters were reviewed primarily by the contributors to this volume, I also thank Roger Azevedo, Max Louwerse, Roger Taylor, and Phil McCarthy, for helping with the review process and Margie Petrowski for helping with the final preparation process of the volume. Finally, I am most grateful to those who contributed their chapters to this volume. Without the work that they have conducted to explore and understand reading strategies, reading strategy interventions, and theories of reading comprehension, this volume most certainly would not have been possible. I thank them for the research they are conducting and for their contributions to this volume.

Who Should Read This Book?

This collection of chapters will be of interest to researchers, educators, and students in the fields of psychology, reading, education, and tutoring technologies. I highly recommend this book to learn more about either reading comprehension or tutoring technologies. It would be particularly appropriate as a resource for a graduate course on reading.

Essentially, this volume will interest anyone who wants to know more about how reading comprehension can improve by using effective, theoretically motivated reading strategies.

—Danielle S. McNamara

I

Theories of Text Comprehension: The Importance of Reading Strategies to Theoretical Foundations of Reading Comprehension

1

An Introduction to Strategic Reading Comprehension

Arthur C. Graesser
University of Memphis

This chapter provides an overview of the conceptual, theoretical, empirical, and pedagogical foundations of reading strategies. It begins by offering a definition and clarification of what it means to have a reading comprehension strategy. The subsequent section contrasts 3 major theoretical frameworks for investigating comprehension in the fields of cognitive science and discourse processing: (a) a construction-integration model, (b) a constructionist theory, and (c) an embodied cognition view. These frameworks offer different claims and commitments with respect to computational architectures and the status of strategies in comprehension. It is recommended that researchers identify the predictions of these and other theoretical frameworks when planning their empirical research on the effectiveness of reading strategies in educational settings. The chapter concludes with a discussion of some challenges that researchers will face when moving from theory to interventions and to assessments of reading comprehension strategies.

Reading is an extraordinary achievement when one considers the number of levels and components that must be mastered. Consider what it takes to read a simple story. The words contain graphemes, phonemes, and morphemes. Sentences have syntactic composition, propositions, and stylistic features. Deep comprehension of the sentences requires the construction of referents of nouns, a discourse focus, presuppositions, and plausible inferences. The reader needs to distinguish given versus new information in the text and implicitly acknowledge what is shared among most readers in a community (called the *common ground*). At more global levels, the reader needs to identify the genre, rhetorical structure, plot, perspective of different characters, narrator, theme, story point, and sometimes the attitude of the author. The coding, interpretation, and construction of all of these levels are effortlessly achieved at a rate of 250 to 400 words per minute by a proficient adult reader.

Comprehension is not always effortless and fast, of course. When beginning readers struggle over individual words, reading is slowed to a near halt and deeper levels of comprehension are seriously compromised. This happens when proficient adult readers struggle with technical expository text on unfamiliar arcane topics, such as a mortgage on a house or the schematics of computer's operating system. Cognitive strategies are particularly important when there is a breakdown at any level of comprehension. A successful reader implements deliberate, conscious, effortful, time-consuming strategies to repair or circumvent a reading component that is not intact. Reading teachers and programs explicitly teach such reading strategies to handle the challenges of reading obstacles. Such strategies are the direct focus of this chapter, and indeed this entire volume.

One could argue that reading strategies are also important for many adults who consider themselves to be skilled readers. There are basically three arguments to bolster this claim. First, many readers do not know whether they are adequately comprehending text. In research on comprehension calibration (Glenberg & Epstein, 1985; Maki, 1998), ratings are collected from readers on how well they believe they have comprehended texts, and these ratings are correlated with objective tests of text comprehension. The comprehension calibration correlations are alarming low (r = .27), even among college students. Acquisition of better reading strategies holds some promise in helping readers improve their comprehension calibration.

Second, many readers have an illusion of comprehension when they read text because they settle for shallow levels of analysis as a criterion for adequate comprehension (Baker, 1985; Otero & Kintsch, 1992). Shallow readers believe they have adequately comprehended text if they can recognize the content words and can understand most of the sentences. However, deep comprehension requires inferences, linking ideas coherently, scrutinizing the validity of claims with a critical stance, and sometimes understanding the motives of authors. Shallow readers believe they are comprehending text when in fact they are missing the majority of contradictions and false claims. Acquisition of better reading strategies is apparently needed to crack the illusion of comprehension in readers who are settling for low standards of comprehension. They need to acquire and implement strategies to facilitate deeper levels of comprehension.

Third, nearly all adults have trouble comprehending technical expository text at deep levels even though they are skilled readers. Deep comprehension of technical text is a difficult challenge, because the reader has minimal knowledge of the technical terms, key conceptualizations, mental models, and other forms of background knowledge. Even those with high relevant background knowledge and general reading skills can struggle. Researchers in my laboratory recently conducted an experiment on students in a college physics course who were assigned to one of three conditions: (a) work on physics problems with an intelligent tutor (called *AutoTutor*), (b) read a textbook on the same content for a duration yoked to the AutoTutor condition, or (c) read nothing (Graesser, Jackson, et al. 2003; Van Lehn et al., in press). Before and after training, there was a pretest and a posttest with multiple-choice questions similar to the Force Concept Inventory (Hestenes, Wells, & Swackhamer, 1992), a test that taps deep physics knowledge. We were thrilled to learn that there were substantial learning gains from AutoTutor, but that is not the main news from the present standpoint. We were surprised to learn that the college students had zero learning gains from reading the textbook, and their posttest scores did not differ from reading nothing at all. A similar finding was obtained on the topic of computer literacy (Graesser, Lu, et al., 2004). Results such as these strongly suggest that the reading strategies of literate adults are far from optimal when considering deep comprehension. Our college students did not achieve deep comprehension on texts about physics and computer literacy even when they had a nontrivial amount of world knowledge on these topics and sufficient reading strategies to land them in college. Acquisition of better strategies of reading comprehension may best be viewed as a lifelong mission.

Some researchers (names intentionally withheld) do not routinely agree that it is worthwhile to teach reading comprehension strategies as an explicit reading objective. Some skeptics argue that the comprehension strategies will follow naturally from reading a large body of texts and from being intrinsically engaged in the content. The problem with this conclusion is that it fails to explain the above findings on comprehension calibration, illusions of comprehension, and the poverty of deep comprehension. Readers are not at all optimally comprehending texts even after decades of practice with reading.

Other skeptics raise the concern that there is a cognitive overhead in applying comprehension strategies and that this overhead can potentially interfere with learning the substantive content. There are two rebuttals to the second worry. Regarding the first rebuttal, a comprehension strategy will have a cognitive cost when first implemented, but these costs will diminish over time as the cognitive strategy becomes more practiced and eventually automatized. As in the case of all skill acquisition, the initial learning requires consciousness, is effortful, is time consuming, and taxes cognitive resources but, after practice, many skills are automatized to the point of being unconscious, effortless, fast, and unimposing in cognitive resources (Ackerman, 1988; La Berge & Samuels, 1974; Perfetti, 1985). Whether the deep comprehension strategies can be completely automatized is at present unanswered in available research, but few would doubt that practice of the strategy will reduce the overload. Regarding the second rebuttal, the reading comprehension strategies I have in mind are intimately connected with substantive content, not detached. The comprehension strategies addressed in this book are sensitive, to varying degrees, to the content expressed in the text and sometimes to the type of subject matter knowledge associated with the text. No one is advocating the use of generic content-free strategies that one often finds in commercial reading programs, such as SQ3R (which stands for Survey, Question, Read, Recite, and Review; Robinson, 1961/1970). The generic strategies of SQ3R are methodically applied to all texts with little or no consideration of the nature of text content. In contrast, the strategies advocated in the chapters of this volume are content sensitive.

The remainder of this chapter is divided into three sections. In the next section, I offer a definition and clarification of what it means to have a reading comprehension strategy. In the section after that, I contrast three major theoretical frameworks for investigating comprehension: (a) a construction-integration model (Kintsch, 1998), (b) the constructionist theory (Graesser, Singer, & Trabasso, 1994), and (c) an embodied cognition view (Glenberg & Robertson, 1999). These frameworks offer different claims and commitments with respect to computational architectures and strategies in comprehension. In the third section, I identify some challenges that researchers will face when moving from theory to interventions and to assessments of reading comprehension strategies.

WHAT IS A READING COMPREHENSION STRATEGY?

A *reading comprehension strategy* is a cognitive or behavioral action that is enacted under particular contextual conditions, with the goal of improving some aspect of comprehension. Consider a very simple-minded strategy for purposes of illustration. Teachers often instruct students to look up a word in a dictionary when they encounter a rare word with which they are unfamiliar. The context would be a word in the text that has low frequency or (more generally) is not in the reader's mental lexicon. The strategic behavioral actions would be to hunt for a dictionary and to locate the word in the dictionary by turning pages. The strategic cognitive actions would be to read the word's definition in the dictionary, to reread the sentence in the text with the word, and then to comprehend the sentence as a whole. One way of specifying this *dictionary-artifact* strategy is with a context-sensitive production rule that has an IF <condition states>, THEN <action sequence> format, such as the rule below.

Dictionary Artifact Strategy

IF <word W is infrequent **OR** Reader does not know meaning of word W>
THEN <(1) reader gets dictionary, (2) reader looks up word W, (3) reader reads dictionary definition, (4) reader rereads sentence with W, and then (5) reader attempts to comprehend sentence as a whole.>

The production rule formalism helps researchers (and potentially teachers) keep track of the details of the strategies and how the strategies get implemented. Failure to heed such detail runs the risk of misapplying the strategies, an occurrence about which researchers and teachers frequently complain. So the reader might apply the rule too often (when the condition elements are not specific enough) or too rarely (when the condition elements are too constrained). A proper tuning of the condition elements and actions is extremely important. The conditional state might be defined either objectively (i.e., the word is rare in the English language) or subjectively (the reader has never encountered the word before). Objective definitions are needed when building some computer technologies, as in the case of a computer tutor that asks the reader whether he or she knows the meaning of low-frequency words. Subjective definitions are needed when training students on self-regulating their application of meta-comprehension strategies (Azevedo & Cromley, 2004; Zimmerman & Schunk, 2001). The point of presenting this production rule is to illustrate the format and context sensitivity of strategies, not to formulate the perfect well-crafted rule.

Most readers are too lazy to hunt for a dictionary every time they encounter a rare word. There also are frequent occasions when the nearest dictionary is miles away. So an alternative strategy is often advocated by reading instructors, namely to “infer the meaning from context.” A *contextual word definition strategy* might be as follows:

Contextual Word Definition Strategy

IF <word W is infrequent **OR** Reader does not know meaning of word W>
THEN <(1) reader rereads previous text for definitional clauses, (2) reader reads subsequent text for definitional clauses, (3) reader rereads sentence with W, and then (4) reader attempts to comprehend sentence as a whole.>

This production rule would have obvious predictions about eye movements because the reader would have regressive eye fixations and forward directed movements in an effort to locate definitional clauses. The strategy influences the cognitive actions of eye movements, whereas there is no need for the behavioral actions of hunting for a dictionary.

There are many other potential strategies involving cognitive actions. For example, readers could be encouraged to assign the unfamiliar word to an ontological category (e.g., an animal) on the basis of context (e.g., X ran through the meadow dodging the trees), even though the reader would not be able to reconstruct the particular subclass or exemplar of the word. Sometimes the text provides enough context to infer that the entity referenced by a word has specific attributes (e.g., it is an animal with stripes that lives in Africa), with enough specification for the reader to continue reading further and glean the major points of the text. Indeed, a good reader knows when it is not worthwhile to fuss with a precise meaning, referent, or attribute specification of a word.

Unfamiliar words can also be handled by nonstrategic mechanisms. For example, many researchers have argued that readers infer the meaning of words from co-occurrences with other words in the large corpus of texts they experience (Anderson, 1990; Landauer & Dumais, 1997). The meanings of words do not normally come from explicit definitions or even from special purpose cognitive strategies during comprehension. Readers ascribe whatever attributes they can to unfamiliar words during reading without their receiving any special-purpose systematic treatment. Accordingly, a strategic treatment of unfamiliar words is a rare or intermittent event rather than the mainstream mechanism. At this point, the jury is still out on the extent to which the treatment of unfamiliar words is handled by strategic versus nonstrategic cognitive processes.

Consider another strategy that has received considerable attention in recent years, namely, the construction of self-explanations during reading (Chi, de Leeuw, Chiu, & LaVancher, 1994; McNamara, 2004; chap. 16, this volume; Millis et al., 2004). When readers build self-explanations, they recruit their world knowledge and personal experiences to make sense out of the explicit text and generate plausible inferences. According to the constructivist theory of text comprehension (Graesser et al. 1994; Magliano, Trabasso, & Graesser, 1999), for example, readers are encouraged to explain the meaning of the text content by generating causes of events, justifications of claims, and other content that explains *why* events in the text occur and *why* the author bothers to mention something. In a story, for example, an action performed by a character should trigger the following *character motive* strategic production rule:

Character Motive Strategy

IF <clause N states that character C performs action A>
THEN <(1) reader retrieves from memory motives that explain A **OR** (2) reader rereads prior text for clauses with motives that explain A **OR** (3) reader constructs inferences from analogous prior experiences with motives that explain A>

Part of the explanations of characters' actions consists of the goals or motives that drive the actions. A character might attack another character for revenge, survival, rescue of a third character, entertainment, and so on. There is ample evidence that deep comprehenders construct more self-explanations (Chi et al., 1994; Trabasso & Magliano, 1996) and that comprehension improves from instructions and training on self-explanations (McNamara, 2004; Pressley et al., 1992). However, researchers have not pinned down the relative timing of self-explanations that come from launching the self-explanation strategy, memory retrieval from text (Component 1 in the preceding example), rereading prior text (Component 2), and generating plausible inferences from prior knowledge (Component 3) when applying the character motive strategy.

Once again, the question arises whether strategies and strategy training is really needed to generate motives that explain the actions of story characters. Perhaps a reader's rich body of experiential knowledge is sufficient to cover the motives of pretty much any action that a character performs in most short stories and novels. World knowledge may come to the rescue very quickly, without the need to deliberately and consciously hunt for motives with the same intensity that some readers do when reading a detective novel that is carefully crafted to disguise character motives. Conscious strategies of self-explanation may be superfluous or disruptive when comprehending actions in simple stories. In contrast, when world knowledge is minimal, such strategies may be particularly important and differentiate shallow versus deep comprehenders. For example, such why-questions and explanations become salient whenever instructions are read in an attempt to assemble furniture or equipment. One important research question is how background knowledge interacts with the acquisition, application, and utility of strategic comprehension strategies (McNamara, 2004; Vitale, Romance, & Dolan, 2006).

It is beyond the scope of this introductory chapter to discuss the many theoretical issues and research questions that merit investigation to advance a scientific understanding of comprehension strategies. However, some of these issues and questions are enumerated in the following list:

1. *What level of representation is being tapped by the strategy?* Strategies differ when pitched at different levels of representation:

word meaning, sentence meaning, local text cohesion, mental models, and global structure versus pragmatic communication.

2. *What prerequisite knowledge or skills are needed to apply the strategy?* For example, jokes are composed with pragmatic content and a rhetorical composition to convey humor, but young children often miss the point of a joke because they lack important wisdom about life or the subtle skills to process the rhetorical level.
3. *What prerequisite knowledge or skills will yield maximal gains from the strategy?* Attempts to connect clauses in a science text in a cohesive manner can be accomplished to the extent there is background knowledge about the science subject matter.
4. *How much training is needed for mastery of the strategy?* A 1- to 2-hr training session is not adequate to master most comprehension strategies. It is not sufficient to memorize verbal articulations of most strategies; it normally takes application and practice on hundreds of texts over many weeks and months.
5. *Does the strategy need to be explicit and conscious, or is unconscious induction adequate?* The question of whether consciousness is required is relevant to the initial acquisition of the strategy as well as the monitoring of a well-practiced strategy.
6. *Does the strategy get executed before, during, or after the mental engagement with the content and subject matter?* The relative timing of strategy execution, apprehension of text content, and recruitment of subject matter knowledge will no doubt attract the attention of researchers for the foreseeable future.
7. *What are the relevant genres and domain knowledge for the strategy?* A *genre* is a category of text, such as a folk tale, a science text, or a persuasive editorial in a newspaper. A strategy that attempts to infer author intent is particularly important for a persuasive editorial and less so, if at all, for a science text. A strategy that attempts to construct a mental image would be important when comprehending a text on assembling equipment, but less so when comprehending a mortgage contract.
8. *Is the strategy best scaffolded by a human or computer?* Some strategies are too subtle and complex to expect a computer system to scaffold. It is too tedious for humans to scaffold strategies that are simple and require thousands of practice trials.

Answers to these questions will vary from strategy to strategy. The hope is that researchers will eventually identify some meta-principles after investigating a large landscape of reading comprehension strategies.

THE STATUS OF STRATEGIES IN DIFFERENT THEORIES OF COMPREHENSION

Discourse psychologists have developed a number of theoretical models of text comprehension during the last two decades. These models make different commitments on the role of comprehension strategies in driving comprehension. It is beyond the scope of this chapter to cover all of the models that have been proposed in recent years. Instead, I contrast three models, each of which serves as a representative of a particular class of models. A construction-integration (CI) model (Kintsch, 1998) will represent a class of bottom-up models, which would also include the *memory-based resonance model* developed by Myers and O'Brien and their colleagues (Myers, O'Brien, Albrecht, & Mason, 1994; O'Brien, Raney, Albrecht, & Rayner, 1997). A constructionist model by Graesser et al. (1994) will represent a class of strategy-driven models, which would also include the *event indexing model* (Zwaan & Radvansky, 1998). An indexical model by Glenberg and Robertson (1999) will represent a class of *embodied cognition models* (Glenberg, 1997; Zwaan, Stanfield, & Yaxley, 2002). Of course, there are a variety of other models that are hybrids, such as the *landscape model* (Van den Broek, Virtue, Everson, Tzeng, & Sung, 2002) and the *Capacity–Constrained Construction-Integration model* (Goldman, Varma, & Cote, 1996). I select these three representative models because they offer rather different perspectives on the role of comprehension strategies in reading.

Construction-Integration Model

Kintsch's (1998) CI model is currently regarded as the most comprehensive model of reading comprehension. Its remarkably simple computational architecture accounts for a large body of psychological data, including reading times, activation of concepts at different phases of comprehension, sentence recognition, text recall, and text summarization. As will soon be apparent, strategies take a back seat in the CI model. Strategies exist, but they do not drive the comprehension engine. Instead, the front seat of comprehension lies in the bottom-up activation of knowledge in long-term memory from textual input (the *construction* phase) and the integration of activated ideas in working memory (the *integration* phase). As each sentence or clause in a text is comprehended, there is a construction phase followed by an integration phase. A *strategy* is simply a piece of knowledge stored in long-term memory that is periodically activated and recruited during integration. It is mixed in the manifold of hundreds or thousands of other concepts, rules, and content during construction and integration. Simply put, strategies are nothing special other than being another set of rules that get activated and integrated.

Like most models in discourse psychology, the CI model assumes that multiple levels of representation get constructed during comprehension. Four of these levels are (a) the surface code, (b) the propositional textbase, (c) the situation model, and (d) the text genre. The *surface code* preserves the exact wording and syntax of the sentences. The *textbase* contains explicit propositions in the text in a stripped-down, logical form that preserves the meaning but not the surface code. The *situation model* (sometimes called the *mental model*) is the referential content or microworld that the text is describing. This would include the people, objects, spatial setting, actions, events, plans, thoughts, and emotions of people and other referential content in a news story, as well as the world knowledge recruited to interpret this contextually specific content. The *text genre* is the type of discourse, such as a news story, a folk tale, or an encyclopedia article. When comprehension succeeds, the representations at all of these levels are harmoniously integrated, yet there is no intentional strategy on the part of the reader to make this happen. It simply falls out naturally from the CI mechanism.

Kintsch's CI model assumes that a connectionist network is iteratively created, modified, and updated during the course of comprehension. As text is read, sentence by sentence (or clause by clause), a set of word concept nodes and proposition nodes are activated (constructed). Some nodes correspond to explicit constituents in the text, whereas others are activated inferentially by world knowledge, rules, and other representations stored in long-term memory. The activation of each node in the network fluctuates systematically during the course of comprehension as each sentence is read. When a sentence (or clause) *S* is read, the set of *N* activated nodes include the explicit and inference nodes affiliated with *S* as well as the nodes that are held over in working memory from the previous sentence *S1* by virtue of meeting some threshold of activation. There are *N* nodes that have varying degrees of activation while comprehending sentence *S*. These *N* nodes are fully connected to each other in a weight space. The set of weights in the resulting *N* × *N connectivity matrix* specifies the extent to which each node activates or inhibits the activation of each of the *N* nodes. The values of the weights in the connectivity matrix are theoretically motivated by the multiple levels of language and discourse. For example, if two proposition nodes (*A* and *B*) are closely related semantically, they would have a high positive weight, whereas if the two propositions contradict each other, they would have a high negative weight.

The dynamic process of comprehending sentence *S* has a two-stage process of construction and integration. During construction, the *N* nodes are activated to varying degrees, specified by an initial activation vector ($\mathbf{a}_1, \mathbf{a}_{2,} \ldots \mathbf{a}_N$). The connectivity matrix then operates on this initial node-activation vector in multiple activation cycles until there is a settling of the node activations to a new final stable activation profile for the *N* nodes. At that point, integration of the nodes has been achieved. Mathematically, this is accomplished by the initial activation vector being multiplied by the same connectivity matrix in multiple iterations until the *N* output vectors of two successive interactions show extremely small differences (signifying a stable settling of the integration phase). Sentences that are more difficult to comprehend would presumably require more cycles to settle.

It is important to emphasize that the mechanisms that drive comprehension are node activations, memory retrieval, integration of nodes in working memory via the connectivity matrix, thresholds for carrying node content across sentences, and other basic mechanisms of memory and cognition. Where do comprehension strategies fit in? A strategy is simply another nodal unit that gets activated, recruited from memory, and incorporated into the connectivity matrix. The generality or specificity of the strategy depends on the history of the texts that have been read, the nature and amount of instructions on the strategy, and the amount of practice in strategy application. A strategy that is taught in a classroom on a particular afternoon would have little or no impact on the reader during subsequent weeks, months, and years. Comprehension strategies have no special status and are not built into the architecture of the CI model in any explicit explanatory fashion.

Constructionist Model

Strategies play a prominent role in the constructionist theoretical framework proposed by Graesser et al. (1994). The distinctive strategies of this model are reflected in its three principal assumptions: (a) reader goals, (b) coherence, and (c) explanation. The *reader goal assumption* states that readers attend to content in the text that addresses the goals of reading the text. When a computer manual is read, for example, it is read very differently when the reader wants to purchase the computer than when the reader wants to fix a broken hard drive. The *coherence assumption* states that readers attempt to construct meaning representations that are coherent at both local and global levels. Therefore, coherence gaps in the text will stimulate the reader to actively think, generate inferences, and reinterpret the text in an effort to fill in, repair, or take note of the coherence gap. The *explanation assumption* states that good comprehenders tend to generate explanations of why events and actions in the text occur, why states exist, and why the author bothers expressing particular ideas. Why-questions encourage analysis of causal mechanisms and justifications of claims. There are other assumptions of the constructionist theory that are shared by many other models, assumptions that address memory stores, levels of representation, world knowledge, activation of nodes, automaticity, and so on, but its signature assumptions address reader goals, coherence, and explanation. The constructionist theory has generated a number of predictions about reading times, inference generation, recall of text information, and summarization; as in the case of the CI model, many of the predictions have been tested and supported, although support for the constructionist model is not as extensive as for the CI model.

The notion that coherence and explanation strategies are the hallmarks of good comprehension places constraints on comprehension. These strategies determine the selection of content that gets encoded, the inferences that are generated, the time spent processing text constituents, and so on. Good readers attempt to bridge incoming sentences with previous text content and with their background knowledge. Good readers are driven by why-questions more than how, when, where, and what-if questions, unless there are special goals to track such information. The explanations of the motives of characters and of the causes of unexpected events in a story are much more important than the spatial position of the characters in a setting, what the character looks like, and the procedures and style of how characters' actions are performed. Such details about space, perceptual attributes, and actions are important when they serve an explanatory function or they address specific reader goals. When readers are asked to monitor why-questions during comprehension, their processing and memory for the text are very similar to normal comprehension without such orienting questions; however, when asked to monitor how-questions and what-happens-next-questions, their processing and memory shows signs of being disrupted (Magliano et al., 1999). Explanations and why-questions are fundamental to the construction of meaning according to the constructionist model. Research on self-explanations, as in the case of Self-Explanation Reading Training (McNamara, 2004) and iSTART (see chap. 16, this volume; McNamara, Levinstein, & Boonthum, 2004), are compatible with this theoretical position, although the precise content that is affiliated with self-explanations is not necessarily restricted to answers to why-questions.

Indexical Hypothesis and Embodiment

Glenberg's *indexical hypothesis* (see chap. 9, this volume; Glenberg & Robertson, 1999) will, for the present purposes, be elevated to the status of a model, because preliminary sketches of a bona fide model are emerging in Glenberg's research program and in Barsalou's (1999) perceptual symbol system. These theoretical positions adopt an embodied theory of language and discourse comprehension. The central theoretical claim is that meaning is grounded in how we use our bodies as we perceive and act in the world. Comprehension of a story is predicted to improve after children have been able to perceive and manipulate the characters and objects in a story scenario. When adults read a manual on assembling a piece of equipment, their comprehension is expected to improve to the extent that they can enact the procedures or at least form visual images of the objects and actions. Readers who have the metacognitive strategy of grounding the entities and events mentioned in the text are expected to show comprehension advantages over those who do not bother taking such extra cognitive steps.

A major point to be made, from the present standpoint, is that the predictions on the effectiveness of strategies on comprehension are dramatically different for the constructionist model and the indexical model. The indexical model would encourage comprehension strategies that involve the construction of mental images of people, objects, spatial layouts, actions, and events expressed in the text. The constructionist model would not encourage these strategies unless they serve the master strategies of building explanations, coherent representations, and representations that address particular reader goals. Indeed, these theoretical models are hardly redundant articulations of the same phenomena with different jargon. Instead, the predictions are decisively different! Perhaps both of the models have some validity, but for different types of texts and comprehension conditions. That is a matter for future research to decide.

CHALLENGES OF MOVING FROM THEORY TO INTERVENTIONS AND ASSESSMENTS OF READING COMPREHENSION STRATEGIES

The contributors to this volume have proposed some reading comprehension strategies that hold some promise in improving comprehension at deeper levels. The strategies and interventions proposed by the contributors to this volume are listed in Table 1.1. The particular strategies and large-scale interventions in this list cover a broad landscape of levels and components at deeper levels. There are strategies designed to improve the comprehension of sentences and local text excerpts; the bridging and connecting of text constituents; the grounding of the text to personal experiences and everyday activities; mastery of the rhetorical structure and genre of text; social interaction with experts, tutors, and peers; processes of question asking, question answering, reflection, and summarization; motivation; and engagement. The community of researchers could hardly be accused of being narrow or paradigm bound.

TABLE 1.1
Strategies and Strategy Interventions

(1)	SERT (Self-explanation Reading Training) and iSTART (Interactive Strategy Training for Active Reading and Thinking) (McNamara, O'Reilly, Rowe, & Levinstein)
(2)	Reciprocal Teaching Method and Questioning the Author (Palincsar, Spiro, and colleagues)
(3)	Concept mapping and a knowledge-focused multi-part reading comprehension strategy (Vitale & Romance)
(4)	PALS: Peer-assisted Learning Strategies (Fuchs & Fuchs)
(5)	CORI: Concept Oriented Reading Instruction (Guthrie, Taboda, & Schuler)
(6)	Text Structure (Williams)
(7)	Structure Strategy tutor (Meyer & Wijekumar)
(8)	Question Asking and Answering (King, Guthrie, Johnson-Glenberg)
(9)	3D Readers (Johnson-Glenberg)
(10)	Joke City (Yuill)
(11)	Indexing and embodiment (Glenberg)
(12)	Summary Street (Caccamise, Franzke, Eckhoff, E. Kintsch, & W. Kintsch)

In this section I identify a number of challenges that this community of researchers will face when they test the impact of the reading strategies on reading improvement. Some challenges can be readily solved with available methods and technologies, but other challenges are far from being handled and will require some radically different approaches to a solution.

Clarifying the Theoretical Predictions

There is ample evidence that comprehension and learning from text is facilitated by a variety of comprehension strategies. Some of these strategies are used by primary school teachers who are known to be effective in teaching reading (Pressley, Rankin, & Yokoi, 1996). Other strategies have not been routinely used by teachers but have been proposed by researchers as being potentially effective. Thus, there is empirical support for claims that comprehension improves by instructions on question asking (King, 1992; Rosenshine, Meister, & Chapman, 1996), reciprocal teaching (Palincsar & Brown, 1984), self-explanation (McNamara, 2004), Concept Oriented Reading Instruction (Guthrie, Wigfield, & Perencevich, 2004), Questioning the Author (Beck, McKeown, Hamilton, & Kucan, 1997), and other strategies advocated by science communities (National Reading Panel, 2000; Snow, 2002).

Nevertheless, it is sometimes difficult to ascertain the theoretical relevance of a particular intervention. Some interventions are compatible with virtually any theory of comprehension, so their value has a practical mission rather than a theoretical mission. For example, all theories would predict benefits from linking text content to personal experiences, so the theoretical status of such a prediction is empty. Some interventions have a "kitchen sink" approach, with a bundle of promising strategies, so it is impossible to pin down which strategy and theoretical prediction is responsible for any significant gain in comprehension. A kitchen sink approach is pragmatically necessary when the researcher runs a serious risk from an ineffective intervention. However, links to theory still end up being murky in kitchen sink interventions. On the flip side, theories are often so subtle and complex that there is no obvious set of intervention conditions to offer practical tests of the theories. Unfortunately, there is an inherent trade-off between pure tests of theoretical predictions and the likelihood that an intervention proves effective.

Sometimes it is unclear what a hypothesis, model, or theory predicts. When advocates of a theoretical position modify their theories or add ad hoc assumptions to accommodate empirical findings, it becomes difficult to reconstruct what really is predicted. To gain some clarity, it is worthwhile to assign each empirical finding or prediction to one of the following four categories: (a) directly articulated in the model, (b) naturally follows from the model but is not directly articulated (which is a virtue of a powerful explanatory model), (c) requires ad hoc assumptions or parameters to accommodate the data or prediction, and (d) impossible to accommodate or out of the scope of the model. A model has greater scope when there is a dominance of categories (a), (b), and (c) and greater decisiveness when there are fewer cells with the value of 3.

To illustrate the proposed analytical scheme, consider the set of predictions in Table 1.2. The left column lists some orienting questions that would either promote deeper comprehension in an intervention for readers who are otherwise shallow comprehenders, elicit answer content on which good comprehenders concentrate if such content is explicitly expressed in the text, or elicit inferences that deep comprehenders routinely generate. For the present purposes, these three techniques will not be differentiated. The numbers in the cells declare the theoretical status according the four-part distinction. The models include the CI model, the constructionist model, and the indexical model covered in the previous section. Also included in Table 1.2 is theoretical framework that inspired the Questioning the Author intervention developed by Beck et al. (1997). This intervention encourages the reader to view the author as a potentially fallible individual who can be questioned about the writing content. So, good comprehenders would query the author with such questions as: Why did the author make a particular statement?, What evidence is there for a claim?, and What is the relevance of an explicit statement to the message as a whole?

Table 1.2 illustrates how disparate the predictions are for the different theoretical models. The CI model does not offer decisive predictions about most of the question categories but could accommodate empirical findings by virtue of ad hoc parameters. Causal explanatory content (answers to "Why did event E occur?") naturally falls out of the connectivity matrix of the CI model (a virtue of a powerful model), for reasons that are not elaborated here, whereas the content reflected in the other categories of questions would require ad hoc assumptions and parameters. The constructionist and indexical models are more decisive and explicit in their claims; there are more cells filled with values of 1 and 4. It is interesting that these two models generate rather different theoretical predictions, which would, I hope, inspire empirical research to see which predictions are confirmed. Questioning the Author also offers predictions that are very different from those of the constructionist and indexing models.

Colleagues might dispute the values presented in these cells of Table 1.2. Indeed, there often are debates over the precise predictions of a particular model, particularly when the models change from publication to publication. The important point to be made here is that tables such as these are valuable in science and educational practice. In the arena of science, they help researchers determine whether a study will help narrow down alternative theoretical positions. In the arena of educational practice, they help researchers select interventions to test, to prepare principled protocols of interventions, and to assign theoretical credit for interventions that work. A scientific framework is increasingly useful to the extent that it motivates intervention conditions that are feasible to implement by teachers, tutors, and technologies.

Grain Size of Strategies

How many strategies should there be? How contextually constrained should a particular strategy be? How specific should one articulate the procedure of applying a strategy? Answers to these questions about grain size are quite different in different fields of inquiry. Researchers in cognitive and discourse psychology would like to see dozens or perhaps hundreds of strategies, each being tuned to appropriate contextual parameters. For example, Strategy S might be appropriate for a particular class of readers (e.g., adults with low subject matter knowledge and general reading ability), text categories (e.g., expository texts on science), and level of representation (e.g., situation model), when later given a particular type of test (e.g., multiple choice). Investigations of higher order interactions among reader, text, task, and representation are advocated by researchers in the area of reading comprehension (McNamara & Kintsch, 1996; Snow, 2002).

TABLE 1.2
Question That Drives Comprehension According to Different Theoretical Positions

Question that drives comprehension	*Construction-integration*	*Constructionist model*	*Indexical, embodiment*	*Questioning the Author*
Why did event E occur?	2	1	3	3
How did event E occur?	3	3	1	3
Why did the author mention event E?	3	1	4	1
What evidence is there that event E is true?	3	4	4	1
What will occur later in the text?	3	3	3	3

(1) directly articulated in the model,
(2) naturally follows from the model, but is not directly articulated,
(3) ad hoc assumptions or parameters are needed to accommodate the data or prediction, and
(4) impossible to accommodate or out of scope of model

Theoretical precision does not necessarily translate well into practice, however. It would be impossible to train a generation of reading teachers how to train children to use hundreds of precisely tuned strategies. They would not have adequate knowledge of cognition, discourse, and language to conduct such detailed training. It is more practical to expect a teacher to implement 5 to 10 strategies that are articulated at a more coarse grain. For example, Self-Explanation Reading Training (McNamara, 2004) has an important strategy called *self-explanation*, which is a covering term for several subtypes of content elaborations that could be specified in detailed analytical theories of explanation. Self-explanation may act as an umbrella term for teachers who apply a number of different concept elaborations, many of which are not among the subtle theoretical distinctions appreciated by scientists. One might wonder what the ideal grain size is for teachers at different points in the educational process. That remains an unanswered empirical question. Perhaps teachers would welcome more subtlety but not the detailed representations expected by a cognitive scientist.

The notion of comprehension strategies has pressed some buttons in the education community because teachers have mechanically applied the strategies. There is liability in having readers apply strategies that are not properly tuned to context. Imagine what the consequences would be if children applied compare/contrast rhetorical structures (see chap. 8, this volume) to every text they read. That would not work well for stories and equipment assembly manuals. Similarly, it would not be adaptive to compose mental images and hierarchical structures for text content unilaterally. These considerations underscore the importance of pitching the grain size at an intermediate level that is not so crude that important distinctions are glossed over but not so refined that the distinctions are misunderstood or ignored by teachers and researchers.

The field of psychometrics can accommodate only three to five theoretical constructs in its assessments of verbal comprehension (see chap. 6, this volume). In the verbal Scholastic Aptitude Test, there are 67 multiple-choice questions and a minimum of 6 to 10 questions per construct. There is considerable discussion of what such constructs should be and how they are grounded in psychological theories (Carroll, 1987; Haladyna, 2004; Mislevy, Steinberg, & Almond, 2003). In chapter 6 of this volume, VanderVeen et al. review their efforts to incorporate cognitive theory into the College Board's verbal Scholastic Aptitude Test. They attempt to identify four to five distinct but related constructs: (a) determining the meaning of words; (b) understanding the content, form, and function of sentences; (c) understanding the situation implied by the text; (d) understanding the content, form, and function of larger sections of text; and (5) analyzing the authors' goals and strategies. These five constructs are approximately aligned with the levels proposed by Kintsch (1998) and by Graesser, Millis, and Zwaan (1997), so some progress has been made in coordinating cognitive theory and psychometric tests.

Cognitive researchers would like to see finer distinctions than five constructs on a psychometric test. Unfortunately, there are properties of the quantitative theories that underlie psychometric tests do not permit it, even if there were hundreds or thousands of test items. The main problem is that the constructs tend to be highly intercorrelated, so it is difficult or impossible to measure the unique contribution of a particular construct. Some discourse researchers are beginning to compose carefully crafted tests that make the constructs orthogonal (Hannon & Daneman, 2001), but unfortunately the tests and tasks are sufficiently unnatural that critics question their representativeness to naturalistic text comprehension. One important question for future research is to develop better tests with naturalistic texts that have cognitive theory aligned with near-orthogonal constructs in psychometric tests. Even when that happens, however, there will probably be limits to the grain size of the constructs. Will there ever be more than five?

Interventions With Humans Versus Computers

Computers are able to train many reading comprehension strategies and are expected to take a more prominent role in the future. Computers do not have the same limitations on fatigue, memory, and grain size that human instructors face. They can potentially diagnose hundreds of reading problems, maintain a student profile on hundreds of variables, tune strategies with an unlimited degree of complexity, and flexibly tailor a particular strategy to the student's learner profile. One could argue that human teachers are not that flexible, but that is an argument that requires empirical investigation.

Critics of computers do not hesitate to point out limitations of computers. Computers are impersonal and lack the vast history of experiences that humans possess and can use at the appropriate times and places. It is noteworthy that these two characteristics can be regarded as strengths in some contexts. For example, some students would rather work with an impersonal computer rather than be embarrassed by their deficits in front of teachers, tutors, and peers. Sometimes the human experiences that teachers share are time consuming and irrelevant to the culture of the learner. Working with the computer is sometimes a better use of the learner's time, especially if it is tailored to the learner's profile to a fine degree. These trade-offs between humans and computers need to be grounded in empirical research to a greater extent than to mere opinion, ideology, and folklore.

Computers are becoming increasingly more sophisticated in providing strategy training. Conventional computer-based training has for decades provided didactic information delivery on descriptions of strategies and examples of strategy use in text, video, and multimedia. However, the advanced learning environments of today are more interactive and adaptive to the abilities of the learner. These include intelligent tutoring systems and trainers that hold conversations in natural language and that have animated conversational agents (Graesser, Lu, et al., 2004; Johnson, 2001). For example, the iSTART system developed by McNamara et al. (2004) uses animated conversational agents to model strategies of experts, to instantiate strategies in peer–agent interactions, to give feedback to learners who try to use the strategies, and to scaffold metacomprehension (analogous to the SERT training by human experts; McNamara, 2004). Modeling–scaffolding–fading techniques have been successfully integrated in many advanced learning environments. The computer systems are substantially more adaptive when they can interpret natural language of users, provide relevant feedback, and advance the interaction in ways that promote learning.

A computer system needs to analyze the activities of the reader if its goals are to be interactive and adaptive. The language contributions of the reader serve as one rich source of reader input that manifests the reader's depth of comprehension. We are fortunate to be at a point in history when computer systems have become very sophisticated in automated analysis of language and discourse. During the last decade, there have been revolutionary advances in computational linguistics (Jurafsky & Martin, 2000) and important advances in discourse processing (Graesser, Gernsbacher, & Goldman, 2003). For example, Coh-Metrix is a computer tool available on the World Wide Web that analyzes texts on multiple levels of cohesion and language (Graesser, McNamara, Louwerse, & Cai, 2004; http://cohmetrix.memphis.edu). Coh-Metrix has the potential to replace standard readability formulas, such as the Flesch–Kincaid Grade Level (Klare, 1974–1975), which rely exclusively on word length and sentence length to scale texts on readability. Coh-Metrix has hundreds of measures of discourse cohesion, syntax, semantics, and word characteristics. Coh-Metrix can potentially be used to select texts for readers to read by intelligent matches to the readers' ability profiles. Coh-Metrix might also be used to analyze verbal contributions of readers when they answer questions or summarize the text.

World knowledge is needed to interpret explicit text and construct plausible inferences. The treatment of world knowledge has traditionally been difficult in computer science, but there have been some breakthroughs in corpus-based statistical algorithms. One notable example of a statistical, corpus-based approach is latent semantic analysis (LSA; Kintsch, 1998; Landauer, McNamara, Dennis, & Kintsch, 2007), which uses a statistical method called *singular value decomposition* to reduce a large Word × Document co-occurrence matrix to approximately 100 to 500 functional dimensions. Each word, sentence, or text ends up being a weighted vector on the *K* dimensions. The match (i.e., similarity in meaning, conceptual relatedness) between two bags of words (single words, sentences, or texts) is computed as a geometric cosine between the two vectors, with values ranging from –1 to 1. LSA-based technology and similar algorithms in computational linguistics are currently being used within a number of applications, such as essay graders that grade essays as reliably as experts in English composition (Burstein, 2003; Landauer, Laham, & Foltz, 2003) and automated tutors that hold conversations in natural language (such as AutoTutor; Graesser, Lu et al., 2004). In this volume, LSA is used in iSTART (chap. 16, this volume), Summary Street (chap. 15, this volume), and systems developed by Magliano and Millis (chap. 5, this volume; Millis et al., 2004).

The prospects of having computers replace human trainers becomes progressively more feasible to the extent that computers become more adaptive to the learner and capable of accurately implementing complex training strategies. Computers are more reliable, more durable, and more capable of accommodating complexity. The systems also have the capacity to train teachers to use some very complex pedagogical strategies. The question of whether computers will replace humans is arguably an empirical one: Can the capacity, complexity, accuracy, cost, and power of automated trainers outstrip what can be supplied by communities of human teachers?

CLOSING COMMENT

These are exciting times for everyone who is attempting to improve reading comprehension and to understand underlying reading mechanisms. We are in the midst of revolutions in educational reform, learning sciences, cognitive sciences, neuroscience, computer science, and information technologies. The need to improve reading literacy in the United States, as well as other countries, is on the radar of the public and government agencies. The role of strategies in improving reading at deeper levels is likely to receive increased attention in the future. This is particularly true in societies that demand more expertise in science, engineering, and technology—areas where world knowledge is modest and the need for comprehension strategies is enormous.

ACKNOWLEDGMENTS

Acknowledgments go to the funders of our research on reading comprehension and discourse processing, including the National Science Foundation (ITR 0325428); the Institute for Education Sciences (IES R3056020018-02, IES R305H050169); and the Department of Defense Multidisciplinary University Research Initiative, administered by the Office of Naval Research under Grant N00014-00-1-0600.

REFERENCES

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*, 288–318.

Anderson, R. C. (1990). Inferences about word meanings. In A. C. Graesser & G. H. Bower (Eds.), *The psychology of learning and motivation: Inferences and text comprehension* (pp. 1–16). New York: Academic.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia. *Journal of Educational Psychology*, *96*, 523–535.

Baker, L. (1985). Differences in standards used by college students to evaluate their comprehension of expository prose. *Reading Research Quarterly*, *20*, 298–313.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.

Beck, I. L., McKeown, M. G., Hamilton, R. L., & Kucan, L. (1997). *Questioning the author: An approach for enhancing student engagement with text*. Newark, DE: International Reading Association.

Burstein, J. (2003). The e-rater® scoring engine: Automated essay scoring with natural language processing. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 113–121). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Carroll, J. B. (1987). Psychometric approaches to cognitive abilities and processes. In S. H. Irvine & S. E. Newstead (Eds.), *Intelligence and cognition: Contemporary frames of reference* (pp. 217–251). Boston: Martinus Nijhoff.

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994) Eliciting self-explanation improves understanding. *Cognitive Science*, *18*, 439–477.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*, 1–19.

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 702–718.

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, *28*, 1–26.

Goldman, S. R., Varma, S., & Cote, N. (1996). Extending capacity-constrained construction integration: Toward "smarter" and flexible models of text comprehension. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 73–114). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Graesser, A. C., Gernsbacher, M. A., & Goldman, S. R. (Eds.). (2003). *Handbook of discourse processes*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Graesser, A. C., Jackson, G. T., Mathews, E. C., Mitchell, H. H., Olney, A., Ventura, M., et al. (2003). Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialogue. In R. Alterman & D. Hirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 1–6). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers*, *36*, 180–193.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, *36*, 193–202.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*, 163–189.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.

Guthrie, J. T., Wigfield, A., & Perencevich, K. C. (Ed.). (2004). *Motivating reading comprehension: Concept-oriented reading instruction*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology*, *93*, 103–128.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher*, *30*, 141–158.

Johnson, W. L. (2001). Pedagogical agent research at CARTE. *AI Magazine*, *22*, 85–94.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice Hall.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal*, *29*, 303–323.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.

Klare, G. R. (1974–1975). Assessing readability. *Reading Research Quarterly*, *10*, 62–102.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 292–323.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Landauer, T. K., Laham, D., & Foltz, P. W. (2003). Automatic scoring and automation of essays with the Intelligent Essay Assessor. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 87–112). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *LSA: A road to meaning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91*, 615–629.

Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*, 1–30.

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–287.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive Strategy Trainer for Active Reading and Thinking. *Behavioral Research Methods, Instruments, and Computers*, *36*, 222–233.

Millis, K. K., Kim, H. J., Todaro, S., Magliano, J., Wiemer-Hastings, K., & McNamara, D.S. (2004). Identifying reading strategies using latent semantic analysis: Comparing semantic benchmarks. *Behavior Research Methods, Instruments, and Computers*, *36*, 213–221.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*, 3–62.

Myers, J. L., O'Brien, E. J., Albrecht, J. E., & Mason, R. A. (1994). Maintaining global coherence during reading. *Journal of Experimental Psychology: Human Learning, Memory, and Cognition*, *20*, 876–886.

National Reading Panel. (2000). T*eaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Pub. No. 00-4769). Jessup, MD: National Institute for Literacy.

O'Brien, E. J., Raney, G. E., Albrecht, J. E., & Rayner, K. (1997). Processes involved in the resolution of explicit anaphors. *Discourse Processes*, *23*, 1–24.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in text: What readers believe versus what they read. *Psychological Science*, *3*, 229–235.

Palinscar, A. S., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*, 117–175.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist, 27,* 91–109.

Pressley, M., Rankin, J., & Yokoi, L. (1996). A survey of instructional practices of primary teachers nominated as effective in promoting literacy. *Elementary School Journal*, *96*, 363–383.

Robinson, F. P. (1970). *Effective study* (4th ed.). New York: Harper & Row. (Original work published 1961)

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, *66*, 181–221.

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.

Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Processes*, *21*, 255–287.

Van den Broek, P., Virtue, S., Everson, M. G., Tzeng, Y., & Sung, Y. (2002). Comprehension and memory of science texts: Inferential processes and the construction of a mental representation. In J. Otero, J. Leon, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 131–154). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (in press). When are tutorial dialogues more effective than reading? *Cognitive Science*.

Vitale, M. R., Romance, N. R., & Dolan, M. F. (2006). A knowledge-based framework for the classroom assessment of student science understanding. In M. McMahon, P. Simmons, R. Summers, D. DeBaets, & F. Crawley (Eds.), *Assessment in science: Practical experiences and educational research* (pp. 1–14). Arlington, VA: National Science Teachers Association.

Zimmerman, B., & Schunk, D. (Eds.). (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Do language comprehenders routinely represent the shapes of objects? *Psychological Science*, *13*, 160–171.

2

Comprehension in Preschool and Early Elementary Children: Skill Development and Strategy Interventions

Panayiota Kendeou, Paul van den Broek, and Mary Jane White
University of Minnesota

Julie Lynch
Saginaw Valley State University

This chapter discusses the development of language comprehension skills in preschool and early elementary school children and the implications of this development for the design of strategy interventions. In the first section, we provide evidence that comprehension starts to develop early in children's lives—before the beginning of systematic instruction—and, further, that the comprehension processes that preschool children use when comprehending events in nonreading contexts are remarkably similar to those used at a later age when reading. In the second section, we provide evidence that comprehension skills and basic language skills (e.g., decoding) develop independently. In the final section, we describe implications of these findings for educational practice. We discuss the importance of developing comprehension strategies at a young age, potential methods for instruction in such strategies, and issues of assessment.

The ability to read and comprehend is critical not only for lifelong learning but also for adequate functioning in society. Despite enormous efforts from researchers, educators, and policymakers to promote reading for all children, many children fail to reach functional levels of literacy (chap. 3, this volume; Paris & Stahl, 2005). Thus, it is critical to understand the nature of young children's early developing language comprehension skills, how they differ from other language skills, and how one can stimulate the development of these skills so that children will be better prepared to excel in reading comprehension when they are formally learning how to read in school.

In this chapter, we focus on the development of language comprehension skills in preschool and then follow the course of comprehension development in children as they become elementary readers. Our discussion is based on the main findings from a longitudinal study we have been conducting with children from ages 4 to 10. From a theoretical point of view, the findings from this study contribute to our understanding of very young children's development of comprehension skills and the developmental path of comprehension skills from prereader to reader. From a practical point of view, the findings from this study have important implications for the design of interventions and strategy instruction that help develop and foster comprehension skills at an early age, well before a child is able to read.

We begin by discussing the development of comprehension skills in preschool and early elementary years as well as the relation between comprehension skills and other basic language skills that are central to reading comprehension. Next, we discuss findings from the longitudinal study we conducted and present a model of reading comprehension development that is supported by the findings of this research. Finally, we discuss educational implications of this model for the design of interventions focusing on the development of comprehension strategy interventions in young children.

DEVELOPMENT OF COMPREHENSION SKILLS

Comprehension means different things to different people. Indeed, comprehension is not a unitary phenomenon but rather a family of skills and activities (Kintsch & Kintsch, 2005; Rapp & van den Broek, 2005; van den Broek et al., 2005). The different types of comprehension share a common core set of processes. A general component in many definitions of comprehension is the interpretation of the information in the text, the use of prior knowledge to interpret this information and, ultimately, the construction of a coherent representation or picture in the reader's mind of what the text is about (e.g., Applebee, 1978; Gernsbacher, 1990; chap. 1, this volume; Graesser & Clark, 1985; Kintsch & van Dijk, 1978; Mandler & Johnson, 1977; Stein & Glenn, 1979; Trabasso, Secco, & van den Broek, 1984). This representation is the foundation from which the reader can retell the story, apply knowledge that has been acquired from the text, identify the theme, and so on.

For adults, the process of understanding written material is automatic in many circumstances, so most of the time we are not aware of the processes we use. During reading, with little or no effort, we identify letters, map letters onto sounds, decode words, understand individual sentences, and make inferences that interconnect different parts of the text (Oakhill & Cain, 2003; van den Broek, 1994). With slightly more effort, we draw connections between the text and our prior knowledge, identify themes, and apply the information we acquire from the text in new situations.

At the core of comprehension is our ability to mentally interconnect different events in the text and form a coherent representation of what the text is about (Trabasso et al., 1984). We can identify different types of connections between events in the texts such as causal, referential, spatial, and so on (Graesser, Singer, & Trabasso, 1994; van den Broek, 1997). Causal connections have received considerable attention in previous research and are believed to be central in the comprehension process (Mandler & Johnson, 1977; Stein & Glenn, 1979; Trabasso et al., 1984). Consider, for example the following set of sentences:

El Niño brings bad times to the coastal cities.
The warm water from excessive rain has less oxygen and nutrients.
Millions of different kinds of fish die.
Thousands of bird varieties starve to death.

When reading this set of sentences, one likely makes a number of causal inferences. Even though is not explicitly stated, one can infer that El Niño entails excessive warm rain. One can also infer that fish die because the warm water has less oxygen. Birds starve to death because fish die and there is not enough food for them. Furthermore, one can also infer that the main cause of all this destruction is El Niño.

When adult readers generate inferences and identify these types of causal connections in the texts they read, they form a mental network representation of the text. Consider, for example, the short story about Tuk the hunter depicted in Figure 2.1. Each sentence in the story is represented in the network as a circle with the corresponding number. The arrows between the circles represent the causal connections between the sentences that the reader may identify.

1. This is the story of a boy named Tuk (took)
2. who lived in the Artic.
3. He wanted to show that he could be brave by hunting for big animals
4. like his father who was a great hunter.
5. Some people do not like idea of hunting.
6. but Tuk's family relies on animals for most of their food and clothing needs.
7. For Tuk's family, hunting is a matter of survival
8. and can be dangerous.
9. Although Tuk was still too young to go on hunting trips
10. and prove how brave he could be,
11. he listened carefully to everything his father told him
12. and was given many of the hunting chores.
13. He helped ready the dogsled for each trip
14. and had learned how to sharpen the hunting spears and knives.

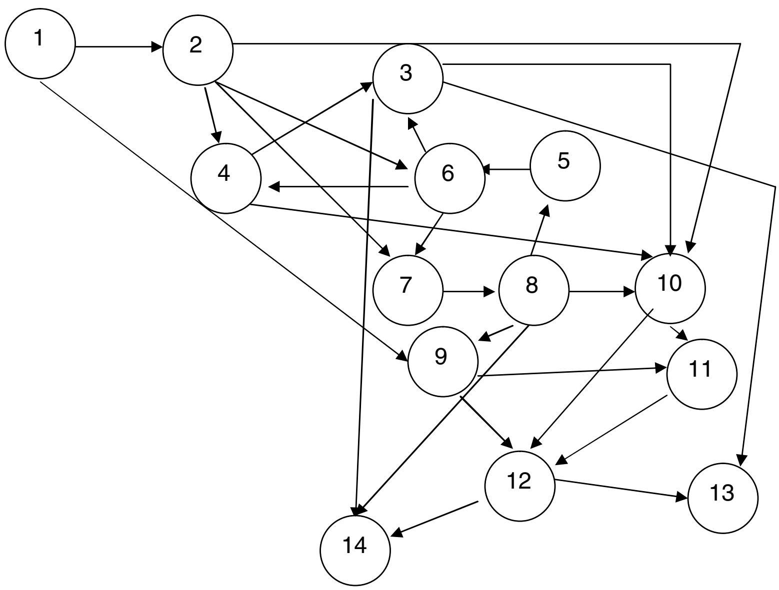


Figure 2.1. Excerpt from "Tuk the Hunter," the written story used in Year 3 of the longitudinal study and the causal network.

For instance, when reading the story we may make the inference that Tuk's desire to become a brave hunter (sentence 3) causes him to want to prove how brave he could be (sentence 10), which in turn causes him to listen carefully to everything his father told him (sentence 11). The network representation in Figure 2.1 is a simple example because the story about Tuk is relatively short and not complicated. Furthermore, we have included only causal connections between sentences in the text and not other potential connections. This simple network, however, allows us to illustrate aspects of adult readers' comprehension.

Previous research suggests that networks such as depicted in Figure 2.1 capture an adult's comprehension of stories (e.g., Goldman & Varnhagen, 1986; O'Brien & Myers, 1987; Trabasso & van den Broek, 1985). Adult readers are more likely to remember events with many causal connections than events with few connections. For example, readers are more likely to remember that Tuk wanted to become a brave hunter (sentence 3, a statement with five connections) or that Tuk wanted to prove how brave he could be (sentence 10, a statement with six connections) than that Tuk helped ready the dogsled (sentence 14, a statement with two connections). Likewise, readers tend to rank the events with many connections as more important than events with fewer connections. Furthermore, readers more often include highly connected events in summaries of the text than events with few connections (van den Broek, 1988). Finally, when readers are asked questions about why an event happened, they respond with answers that are based on the causal connections in the text (Trabasso, van den Broek, & Liu, 1988). For example, when asked why Tuk was too young to go on a hunting trip (sentence 9), readers likely answer with the connected statement that it is because hunting is dangerous (sentence 8).

For adults, the process of making connections often is automatic, but for many young children it is effortful (as it is for adults when the materials are challenging). The ability to make connections in young children develops gradually over time (e.g., Bourg, Bauer, & van den Broek, 1997; Goldman & Varnhagen, 1986). Previous research has revealed developmental trends with respect to three aspects of comprehension: (a) general sensitivity to the causal structure of the narrative; (b) a focus on internal events, such as protagonists' goals in causal structures, as opposed to a focus on external events, such as actions; and (c) the inclusion of between-episode connections in the mental representation of the text as opposed to only within-episode connections (for a review, see van den Broek, 1997).

With regard to the first aspect, the causal structure of the narrative has been found to guide even young children's comprehension of narratives, but it does so less strongly than for older children and adults. In contrast, many nonstructural properties of events, including superficial ones, such as the event's vividness, affect young children's attention. With development, the role of structural properties of narratives increases, whereas that of nonstructural properties of narratives decreases (van den Broek, 1997). With regard to the second aspect, evidence suggests that children's understanding of characters' goals is more sophisticated than previously thought (Goldman & Varnhagen, 1986), but young children do have a tendency to focus much of their attention on observable, concrete actions rather than on internal causes such as characters' goals. For example, children watching a cartoon with a humorous scene where the characters chase each other are more likely to attend to that scene than to scenes that support the development of the structure of the story. Finally, with regard to the third aspect, very young children have a tendency to limit their connection-building to events within each episode and fail to identify possible connections between episodes (Trabasso & Nickels, 1992). Yet the between-episode connections are most likely to be related to the overall theme or message of the text and hence are central to obtaining a complete picture of the meaning of the narrative as a whole. As children grow older, their abilities in all three of these aspects develop and improve (Bourg et al., 1997; Trabasso et al., 1984; van den Broek, 1989a).

Conclusions regarding the development of children's comprehension are based on studies that have primarily involved elementary school children. Research on younger children (e.g., preschool, kindergarten) is limited because children at these ages cannot yet read, and therefore it is not possible to assess their reading comprehension skills. Comprehension can be assessed, however, in a nonreading context by presenting stories in other media. For example, stories can be presented using pictures (Paris & Paris, 2003), aurally (Diakidoy, Stylianou, Karefillidou, & Papageorgiou, 2005; Palincsar, 1991), or via television (Lorch & Sanchez, 1997; van den Broek, Lorch, & Thurlow, 1996).

When using different media to assess comprehension, the underlying assumption is that comprehension skills transfer across media. This positive transfer of skills is plausible for several reasons. First, the structural story factors that predict which events children will recall when they are presented with a story are not medium specific. For instance, narratives that are spoken, televised, or written follow similar structural patterns, such as having a goal structure (i.e., episodes that have a goal–action–outcome structure) and deriving coherence from causal and other connections. Second, children's ability to make inferences follows a developmental pattern that is consistent across different media. For instance, in narratives presented in different media, preschool children begin by making concrete local inferences, and by early elementary school they are able to make abstract global inferences (Goldman & Varnhagen, 1986; van den Broek, 1989b).

EXAMINING COMPREHENSION SKILLS

We have directly tested the idea that comprehension skills generalize across different media in our own research (Kendeou et al., 2005; Lynch et al., 2006; van den Broek et al., 2005). In a longitudinal study, we investigated the development of comprehension skills in different media, their relation to other basic language skills, and their contribution to later reading comprehension. In this study, two cohorts of children ages 4 ($N = 113$) and 6 ($N = 116$) were presented with aural and televised stories and were asked to demonstrate their comprehension of the stories through recall and by answering factual and inferential questions about selected events. *Factual questions* refer to specific events in the stories, whereas *inferential questions* refer to events that can be deduced from events in the stories. Every 2 years, this procedure was repeated, and when children were at an age when they could read (i.e., 8 and 10 years old) they were also presented with written stories and were asked to read, recall, and answer questions about those stories.

The findings show that within each age group (4-, 6-, 8- , and 10-year-olds) aural and television comprehension were highly interrelated. When reading comprehension was included (for the older children), all three comprehension measures were highly interrelated. For all age groups, comprehension skills in different media were not systematically related to basic language skills such as phonological awareness and letter and word identification. Vocabulary scores were related to both comprehension and basic language skills.

We further explored the relations between the different variables within each age group using factor analysis, a multivariate data reduction technique that allows one to detect interrelations among large number of variables and form separate groups of variables based on those interrelations (i.e., a factor). At the first time point, an unforced factor analysis yielded three factors for both the 4-year-old and the 6-year-old children. The first factor included overall recall and recall of highly causally connected events across different media. The second factor included performance on factual and inferential comprehension questions across different media plus vocabulary. The final factor included basic language skills, such as phonological awareness and letter and word identification. These three factors were the same for both 4- and 6-year-old children.

At the second time point, when the children were now 6 and 8 years old, respectively, a second unforced factor analysis was again conducted within each age group. For the 6-year-old children, three factors emerged that were similar to those identified for the 6-year-old children at the first time point. For the 8-year-old children, two factors emerged. The first factor for the

8-year-old children included overall recall and recall of highly causally connected events across media (audio, audiovisual, and written text). The second factor included basic language skills, vocabulary, and comprehension questions across media. However, overall recall and recall of highly causally connected events for the written text were also strongly related to the second factor, suggesting an integration of basic skills and comprehension once children were able to read.

The findings from this exploratory factor analysis allow us to draw a number of conclusions with respect to young children's comprehension processes. The first conclusion relates to the consistency of children's recall of stories across media. Children's recall across media loaded as a separate factor at ages 4, 6, and 8, suggesting that there are underlying similarities in the processes by which children recall information from aural, televised, and written narratives. Furthermore, these processes are relatively separate from basic language skills and vocabulary. The second conclusion relates to the consistency of children's answers to comprehension questions across different media. Children's answers to factual and inferential questions loaded separately from recall at ages 4, 6, and 8, suggesting that question comprehension processes differ from recall comprehension processes. Furthermore, question comprehension processes relate to children's vocabulary at ages 4 and 6 and to vocabulary and other basic language skills at age 8. The third conclusion relates to the pattern observed when reading entered the equation at age 8. Children's recall of written text loaded on both factors that emerged, suggesting that recall comprehension of written texts is related to recall processes across media, vocabulary, and basic language skills.[1]

These conclusions raise the question of whether early comprehension processes, which are similar across media and develop separately from basic language skills, predict later narrative comprehension processes. To address this question, we conducted a series of hierarchical regression analyses for each cohort of children. In the first cohort, this analysis demonstrated that, in preschool children, early narrative comprehension processes at age 4 predicted later narrative comprehension processes at age 6 across audio and television media. In the second cohort, a similar analysis demonstrated that narrative comprehension processes at age 6 predicted later narrative comprehension at age 8 across audio and television media. The findings from this set of analyses suggest that comprehension skills that are developed in preschool are important and contribute to children's comprehension skills as they enter elementary school.

[1]Similar analyses are being conducted on data from these children returning at the third time point, when they were 8 and 10 years old. Preliminary analysis has indicated findings similar to those for the 8-year-old children.

Moreover, these early comprehension processes predict later reading comprehension, and they do so separately from basic language skills and vocabulary. We conducted a hierarchical regression analysis in which reading comprehension at age 8 was entered as the dependent variable and children's narrative comprehension processes in nonwritten media, basic skills, and vocabulary at age 6 were entered as the independent variables. The results showed that early narrative comprehension processes across media at age 6 predicted reading comprehension processes at age 8 over and above basic language skills and vocabulary.

New View of Reading Comprehension: Integration of Findings

These findings allow us to draw several conclusions. First, narrative comprehension processes are remarkably similar across media and develop in parallel with basic language skills well before children begin to read. Second, early-developing comprehension processes join with basic language skills in affecting reading comprehension when the child becomes a reader. These conclusions present a view of development in reading comprehension that differs from the commonly held view that decoding processes develop before comprehension processes develop.

We tested this new view of development in children as they developed from kindergarten (age 6) to second grade (age 8). Path modeling showed that comprehension in different media and word identification at age 6 significantly and independently predicted reading comprehension at age 8. Although vocabulary at age 6 did not predict reading comprehension at age 8 directly, it related to reading comprehension indirectly, because it predicted both comprehension in different media and word identification at age 6.

The new view emphasizes (see Fig. 2.2) that both comprehension and basic language skills contribute to reading comprehension. Although the bulk of attention in recent years has been on fostering basic language skills and vocabulary, comprehension skills are equally important for the development of reading comprehension. Thus, it may be beneficial to foster comprehension skills early on in children's lives.

EDUCATIONAL IMPLICATIONS

Our findings indicate that comprehension skills are an important component of reading comprehension and that they are independent of other, basic language skills typically associated with early reading. Moreover, the fact that comprehension skills in the preschool years predicted later reading comprehension—again,

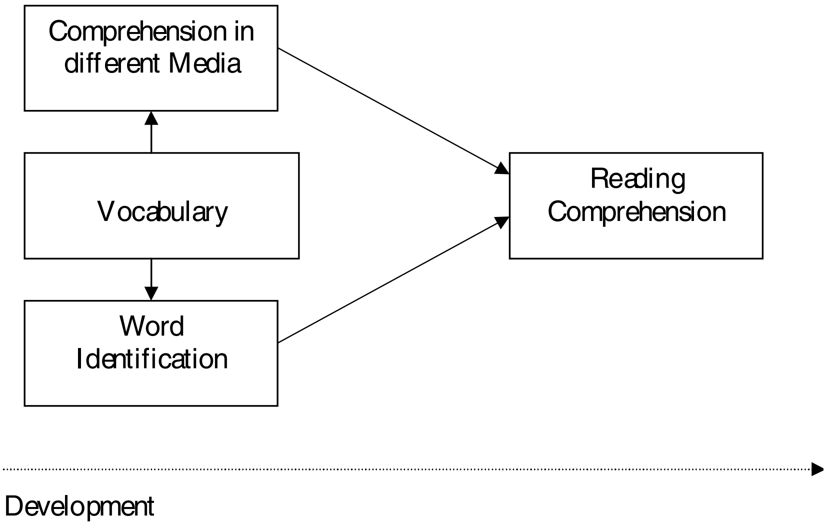


Figure 2.2. The new view of reading comprehension development.

independently from the other basic skills—suggests that the development of these skills may benefit from separate and targeted instruction, as early as the preschool years.[2] Because these children are not yet proficient at reading, such instruction would need to take place in nonreading contexts. In this section, we explore various ways in which comprehension skills and strategies can be encouraged at an early age. These methods—and others like them—are promising strategy interventions and worthy of direct tests of their effectiveness.

Using Different Media to Foster Comprehension Skills

One implication for strategy intervention relates to the types of materials used to develop or improve comprehension skills in young children. Our research

[2]In this chapter we emphasize the similarity in comprehension across media and the potential this offers for instruction. It is important to note, though, that although the similarities between comprehension in different media are large, comprehension is not identical across media. Indeed, there are aspects of texts—for example, layout, text structure, serial presentation—that are unique to reading comprehension. Therefore, it would be an error to conclude that "reading comprehension" simply reduces to "language comprehension."

suggests that comprehension strategies generalize to a considerable degree across media and, therefore, that comprehension interventions and instruction can be conducted using nonreading materials—for example, in the form of aural or televised stories (van den Broek et al., 1996). Indeed, there already is evidence that aural stories can be used effectively in strategy instruction for older children, at elementary school age. In Reciprocal Teaching, Palincsar and Brown (1984) improved reading comprehension skills of elementary school children with reading difficulties by using aurally presented texts. In this program, children are taught to effectively use four different strategies: (a) prediction, (b) clarification, (c) question generation, and (d) summarization. Also, in Peer Assisted Learning Strategies (PALS), Fuchs, Fuchs, Mathes, and Simmons (1997) used aurally presented material for the development of reading comprehension skills in elementary school students with special needs. In this program, teachers pair students in the class so that partners work on different activities that address the problems they are experiencing (e.g., letter names, letter sounds, first-sound identification, phonological awareness, comprehension, etc.). Our research helps understand why these approaches are effective and which components are most likely to contribute to their effectiveness. Moreover, it extends such approaches in two important ways: (a) by indicating that similar strategies for comprehension skill development may be successful at even younger ages, at the preschool level (cf. chap. 7, this volume) and (b) by suggesting that televised stories may be used in a similar manner.

Although much maligned, television can be an appealing medium for fostering comprehension especially in very young children. The visual nature of television stories is motivating to young children, as is the often-positive manner in which a child's everyday experience is reflected in age-appropriate shows such as Sesame Street and Blues Clues. In addition, television viewing can easily be used with many children at a time and can take place inside and outside of schools. Most important, from a comprehension-fostering point of view, television provides a unique opportunity for children to learn and be taught comprehension strategies that are not completely dependent on verbal skills. The context of television can be a focal point for teachers or parents to interact with their students or children around story lines, to model and reinforce coherence-building strategies, including, for example, the generation of explanatory and predictive inferences. It is important to note that simply passively watching television is insufficient for developing comprehension strategies in young children. It is through the systematic application and practice in strategy use (e.g., via systematic inference-based questions; see below) that television stories are most likely to be effective tools.

Assessing Comprehension Skills

A second, related implication pertains to the assessment of comprehension skills. Just as television and other nonreading stories can be used to foster comprehension skills, they can be used to assess those skills. The findings of our research suggest three principles for assessing comprehension (van den Broek et al., 2005). First, comprehension skills develop early in a child's life: Even 4-year-old children show the hallmarks of comprehension that define comprehension in much older, proficient readers. Second, comprehension profiles in young children for different media are highly related, sharing a common core. Third, the assessment of such comprehension skills should focus on multiple aspects of comprehension rather than a single aspect. In addition, it should include both quantity and quality of comprehension. For example, in our research children watch television narratives and are asked to recall everything they remember from the stories. In addition, they answer questions aimed at different levels of inference-making related to the causal structure of the narrative. Quantity of comprehension is indicated by the amount of information children remember and how many questions they answer correctly, whereas quality is indicated by the extent to which their recall focuses on the events and facts that have many connections to other events and facts (and, hence, are central to the structure of the text), and to the extent to which they answer comprehension questions at high levels of inference-making. As children's skills develop, quantitative aspects of comprehension will improve but, more important, so will the qualitative aspects.

A method for assessing comprehension skills in young children such as this is likely to be useful not only for researchers but also for teachers to determine whether their strategy interventions are effective and how individual children progress. Indeed, effective teachers ask appropriate questions to determine how deeply children understand a narrative or episode (Pressley & McCormick, 1995). These questions are similar to the inferential questions we created in our studies. These teachers know that the quality of the understanding is at least as important as is the quantity of facts that are retained.

Strategy Interventions

This leads to a third implication from our research, concerning the types of strategy interventions that might be used for young children. Our results suggest that many strategy-fostering activities that have been found to be effective with older children can be adapted "downward" to children of preschool age. There are ample examples in the literature with respect to the types of strategy interventions one may use to develop and foster comprehension skills in young children and adults (e.g., Beck, McKeown, Sandora, Kucan, & Worthy, 1996; Chi, deLeeuw, Chiu, & LaVancher, 1994; Fuchs et al., 1997; McNamara, 2004; Palincsar & Brown, 1984; Sinatra, Beck, & McKeown, 1993; Yuill & Oakhill, 1988). Most of these interventions involve questioning. For example, in Reciprocal Teaching, elementary school children learn to self-question, in PALS children learn to address different questions in their pairs, in the Questioning the Author program children also learn to generate questions, and so on. Similar interventions can be used with younger children after viewing a television story.

Our findings suggest that comprehension-strategy interventions are likely to be particularly effective if they are based on the causal–logical structure of the text. Instead of just teaching children to question, teachers might structure the questioning activities to focus on the important events as indicated by their causal role in the narrative. Such questions can help children identify causes (e.g., "Why did this happen?") and consequences (e.g., "What do you expect to happen next?"), characters' goals (e.g., "What made him/her do that?"), character's actions (e.g., "What did he/she do?") and themes (e.g., "What did the character learn?"). The selection of questions would be based on the causal network structure of the story and would focus on central events. These more directed interventions could help preschool children develop the skills they need to identify and infer meaningful connections (e.g., causal ones) between various parts of the story when such connections are needed for comprehension. Ultimately, the aim would be for the children to internalize and transfer these strategies to novel situations. These interventions have potential not only for young children but for children of all ages.

Thoughtful Versus Thoughtless Strategy Instruction

In developing comprehension strategy interventions, it is tempting to assume that struggling readers' comprehension can be improved by teaching them to use strategies that good or experienced readers have developed and use. Indeed, this assumption is—implicitly or explicitly—the starting point for many interventions (in reading as well as in other areas). However, it also happens to be an assumption that is naïve and simplistic (at best) or erroneous and counterproductive (at worst). This is because it ignores the fact there may be structural reasons for the struggling readers' failure not to engage in the proposed strategic behaviors. The basis for the assumption is that, by definition, strategies are conscious and controllable activities (Pressley, Forrest-Pressley, Elliott-Faust, & Miller, 1985) and, furthermore, that they are thought to enhance cognitive processing by, in the case of reading, facilitating the understanding and memory of the textual material (Pressley & McCormick, 1995). For example, when the information in the text warrants it, good readers often make predictive inferences about what will happen next in the narrative. So, should we teach struggling readers to make predictive inferences also? No, because it is plausible that they do not spontaneously generate predictive inferences because all their cognitive resources (e.g., working memory) are going to more essential processes, such as making inferences that explain what they currently are reading. Teaching these readers to make predictive inferences—without somehow allowing them to muster up the extra needed resources—would siphon resources from the generation of explanatory inferences. The result could be decreased performance. A second example concerns the use of self-questioning. Good readers often have developed the strategy to self-question during reading. This strategy helps them gain a deep understanding of the text, remember the text better, and generate inferences that are central for coherence. Again, teaching struggling readers to self-question may or may not improve their comprehension, depending on whether doing so interferes with even more essential processes. Thus, if struggling readers' difficulty lies primarily in decoding or interpreting the sentence itself, then the resources required for these basic processes would leave none for self-questioning. In fact, in such a case questioning during reading may interfere with whatever comprehension struggling or beginning readers can muster spontaneously (van den Broek, Tzeng, Risden, Trabasso, & Basche, 2001).

Knowing that struggling readers may have trouble with the application of certain strategies that seem to come naturally to good readers can be useful in developing strategy interventions, but only if this is coupled with an understanding of the underlying causes for such difficulty. Identifying the underlying causes of struggling readers' difficulties, though, is not a trivial task. For instance, the causes may concern reader characteristics, such as basic skills, prior knowledge, attention, motivation, and so on (van den Broek & Kremer, 1999) or text properties such as text structure (e.g., Taylor & Williams, 1983; chap. 8, this volume); or the instructional context, such as explicitness of instruction (e.g., Darch & Kameenui, 1987). However, research in cognition, education, and psychology provide support for the following three principles, which offer suggestions about how we can work systematically in the identification and remediation of readers' difficulties:

1. It is important to gain an understanding of both the processes in which struggling readers engage as they read and the outcomes of such processes to develop effective reading comprehension interventions.
2. Readers at different grade and performance levels are likely to differ in the cognitive processes they recruit to read. Thus, it is important to identify the unique patterns of processes or profile for each subgroup.

3. The difficulties that struggling readers have may manifest themselves differently in different types of texts. Thus, it is important to explore whether strategy interventions generalize across different types of text.

By considering factors such as reader characteristics (e.g., basic skills, background knowledge, inferential skills, attention-allocation skills, working memory capacity, etc.), text properties (e.g., structure of texts), and instructional context (e.g., peer-assisted learning, group learning, and direct instruction)—factors that have been demonstrated to determine successful comprehension (e.g., van den Broek & Kremer, 1999)—we may be able to determine the origin of the struggling readers' difficulties and to design or select appropriate strategy interventions for reading comprehension rather than blindly assume that teaching them good readers' strategies will solve their problems.

CONCLUSION

Children's comprehension is remarkably similar to and yet systematically different from that of adult readers. Even preschool children interconnect events in the stories they hear or see, generate inferences, draw on their background knowledge, and ultimately construct a coherent, mental network representation of the narratives—much like older children and adults do. However, children's ability to accurately create this network and identify different types of connections develops with age. As children grow older, they acquire sensitivity to the causal structure of a narrative, the ability to focus on internal events of characters, and the ability to make connections between episodes and across texts, in addition to connections within episodes of one text.

The research reviewed in this chapter shows that preschool and elementary school children's comprehension skills are remarkably similar across different media and develop in parallel with basic language skills. Moreover, both early narrative comprehension skills and basic language skills begin to develop prior to reading, and both uniquely predict reading success when the child enters elementary school.

The findings of our research suggest that, although much attention in preschool language programs has been on fostering basic language skills that support decoding, development of and instruction in comprehension skills are at least as important. The fact that such comprehension skills predict later reading comprehension over and above other basic language skills suggests that fostering comprehension skills in preschool children may help them become better comprehenders when they begin to read years later.

The literature on reading instruction in the school years provides a wide array of strategy interventions that address potential causes of reading difficulties (e.g., Dole, Duffy, Roehler, & Pearson, 1991; Fuchs et al., 1997; chap. 11, this volume; Palincsar & Brown, 1984; Pressley, 1998, 2000; Yuill & Oakhill, 1988) and that have been shown to be effective in elementary school instruction. These interventions include strategy instruction, questioning, and monitoring. These existing strategy interventions can be adapted—based on an understanding of what components are most likely to be responsible for their success —and new ones specifically developed for preschool children. In particular, the use of nonreading contexts such as television viewing/listening and of questioning activities that are closely connected to the causal structure of the narratives are promising avenues for the development of comprehension strategy interventions in preschool children. These and other adaptations are worth testing in applied studies (see chap. 7, this volume, for an initial attempt to adapt PALS for first-grade students).

We have argued that it is essential to select or design interventions that address the underlying causes of the difficulties struggling readers face rather than falling victim to the "teach poor readers to behave like good readers" fallacy. Effective interventions will affect readers' actual processes during comprehension, particularly at points where readers' comprehension processes break down. In doing so, they will also affect the product of reading, the memory representation and understanding that our students have after they completed reading a text.

ACKNOWLEDGMENTS

This research was supported by funding from the Center for the Improvement of Early Reading Achievement/National Science Foundation, Grant TR R305 R7004; the Guy Bond Endowment in Reading and Literacy; a Golestan fellowship from the Netherlands Institute for Advanced Study in the Humanities and Social Sciences; the National Institute of Child Health and Human Development (HD-07151); and by support from the Center for Early Education and Development, the Eva O. Miller Graduate School Fellowship, and the Center for Cognitive Sciences, all at the University of Minnesota.

REFERENCES

Applebee, A. N. (1978). *The child's concept of a story: Ages two to seventeen*. Chicago: University of Chicago Press.

Beck, I. L., McKeown, M. G., Sandora, C., Kucan, L., & Worthy, J. (1996). Questioning the author: A yearlong classroom implementation to engage students with text. *Elementary School Journal*, *96*, 385–414.

Bourg, T., Bauer, P., & van den Broek, P. (1997). Building the bridges: The development of event comprehension and representation. In P. van den Broek, P. Bauer, & T. Bourg (Eds.), *Developmental spans in event comprehension and representation: Bridging fictional and actual events* (pp. 385–407). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chi, M., deLeeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439–477.

Darch, C. B., & Kameenui, E. J. (1987). Teaching LD students critical reading skills: A systematic replication. *Learning Disability Quarterly*, *10*, 82–91.

Diakidoy, I. N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2005). The relationship between listening and reading comprehension on different types of texts at increasing grade levels. *Reading Psychology*, *26*, 55–80.

Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). Moving from the old to the new: Research on reading comprehension instruction. *Review of Educational Research*, *61*, 239–264.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, *34*, 174–206.

Gernsbacher, M. A. (1990). *Language comprehension as a structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Goldman, S. R., & Varnhagen, C. K. (1986). Memory for embedded and sequential story structures. *Journal of Memory and Language*, *25*, 401–418.

Graesser, A. C., & Clark, L. F. (1985). *The structures and procedures of implicit knowledge*. Norwood, NJ: Ablex.

Graesser, A., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative comprehension. *Psychological Review*, *101*, 371–395.

Kendeou, P., Lynch, J. S., van den Broek, P., Espin, C., White, M., & Kremer, K. E. (2005). Developing successful readers: Building early narrative comprehension skills through television viewing and listening. *Early Childhood Education Journal, 33,* 91–98.

Kintsch, W., & Kintsch, E. (2005). Comprehension. In S. Stahl & S. Paris (Eds.), *Children's reading comprehension and assessment* (pp. 71–92). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kintsch, W., & van Dijk, T. A. (1978). Towards a model of text comprehension and production. *Psychological Review*, *85*, 363–394.

Lorch, E. P., & Sanchez, R. P. (1997). Children's memory for televised events. In P. van den Broek, P. J. Bauer, & T. Bourg (Eds.), *Developmental spans in event comprehension and representation* (pp. 271–291). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lynch, J. S., van den Broek, P., Kremer, K. E., Kendeou, P., White, M., & Lorch, E. P. (2006). *The development of narrative comprehension in preschool and kindergarten: Evidence from complex, authentic measures*. Manuscript submitted for publication.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111–151.

McNamara, D. S. (2004). SERT: Self-Explanation Reading Training. *Discourse Processes*, *38*, 1–30.

Oakhill, J. V., & Cain, K. (2003). The development of comprehension skills. In T. Nunes & P. Bryant (Eds.), *Handbook of children's literacy* (pp. 155–180). Dordrecht, The Netherlands: Kluwer Academic.

O'Brien, E. J., & Myers, J. L. (1987). The role of causal connections in the retrieval of text. *Memory & Cognition*, *15*, 419–427.

Palincsar, A. S. (1991). Scaffolded instruction of listening comprehension with first graders at risk for academic difficulty. In A. McKeough & J. L. Lupart (Eds.), *Toward*

the practice of theory-based instruction: Current cognitive theories and their educational promise (pp. 50–65). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction*, *1*, 117–175.
Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, *38*, 36–76.
Paris, S. G., & Stahl, S. A. (2005). *Children's reading comprehension and assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Pressley, M. (1998). Comprehension strategies instruction. In J. Osborn & F. Lehr (Eds.), *Literacy for all: Issues in teaching and learning* (pp. 113–133). New York: Guilford.
Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. E. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 545–561). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Pressley, M., Forrest-Pressley, D., Elliott-Faust, D. L., & Miller, G. E. (1985). Children's use of cognitive strategies, how to teach strategies, and what to do if they can't be taught. In M. Pressley & C. J. Brainerd (Eds.), *Cognitive learning and memory in children* (pp. 1–47). New York: Springer-Verlag.
Pressley, M., & McCormick, C. B. (1995). *Advanced educational psychology for educators, researchers, and policymakers*. New York: HarperCollins.
Rapp, D. N., & van den Broek, P. (2005). Dynamic text comprehension: An integrative view of reading. *Current Directions in Psychological Science*, *14*, 276–279.
Sinatra, G. M., Beck, I., & McKeown, M. (1993). How knowledge influenced two interventions designed to improve comprehension. *Reading Psychology*, *14*, 141–163.
Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (Vol 2, pp. 53–120). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Taylor, M. B., & Williams, J. P. (1983). Comprehension of learning-disabled readers: Task and text variations. *Journal of Educational Psychology*, *75*, 743–751.
Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, *15*, 249–275.
Trabasso, T., Secco, T., & van den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*, 612–630.
Trabasso, T., van den Broek, P., & Liu, L. (1988). A model for generating questions that assess and promote comprehension. *Questioning Exchange*, *2*, 25–38.
van den Broek, P. W. (1988). The effects of causal relations and hierarchical position on the importance of story statements. *Journal of Memory and Language*, *27*, 1–22.
van den Broek, P. W. (1989a). Causal reasoning and inference making in judging the importance of story statements. *Child Development*, *60*, 286–297.
van den Broek, P. W. (1989b). The effects of causal structure on the comprehension of narratives: Implications for education. *Reading Psychology*, *10*, 19–44.
van den Broek, P. (1994). Comprehension and memory of narrative texts. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 539–588). London: Academic Press.
van den Broek, P. (1997). Discovering the cements of the universe: The development of event comprehension from childhood to adulthood. In P. van den Broek, P. Bauer, & T. Bourg (Eds.), *Developmental spans in event comprehension: Bridging fictional and actual events* (pp. 321–342). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

van den Broek, P., Kendeou, P., Kremer, K., Lynch, J. S., Butler, J., White, M. J., & Lorch, E. P. (2005). Assessment of comprehension abilities in young children. In S. Stahl & S. Paris (Eds.), *Children's reading comprehension and assessment* (pp. 107–130). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

van den Broek, P., & Kremer, K. E. (1999). The mind in action: What it means to comprehend during reading. In B. Taylor, M. Graves, & P. van den Broek (Eds.), *Reading for meaning* (pp. 1–31). New York: Teachers College Press.

van den Broek, P., Lorch, E. P., & Thurlow, R. (1996). Children's and adult's memory for television stories: The role of causal factors, story-grammar categories, and hierarchical level. *Child Development*, *67*, 3010–3028.

van den Broek, P., Tzeng, Y., Risden, K., Trabasso, T., & Basche, P. (2001). Inferential questioning: Effects on comprehension of narrative texts as a function of grade and timing. *Journal of Educational Psychology*, *93*, 521–529.

Yuill, N., & Oakhill, J. (1988). Effects of organizational cues and strategies on good and poor comprehenders' story understanding. *Journal of Educational Psychology*, *80*, 152–158.

3

Issues of Causality in Children's Reading Comprehension

Jane Oakhill
University of Sussex

Kate Cain
Lancaster University

This chapter focuses on the skills and abilities that are theoretically relevant to reading comprehension skill and that have been found to discriminate consistently between good and poor comprehenders. We explore the relative contribution of these skills to the prediction of text comprehension skill (as opposed to word reading) during the period when children move from being beginner to fluent readers (ages 7–11) and present evidence of which skills and abilities have a causal influence on the development of reading comprehension.

Our previous work, and that of other researchers, has provided information about the skills and abilities that are associated with reading comprehension in children (for a summary, see Cain & Oakhill, 2004). To become a successful reader, a child must be able to decode the individual words on the page and must be able to comprehend the text. Word reading and reading comprehension are highly related skills: Correlations between them fall within the range of .30 to .77 (Juel, Griffith, & Gough, 1986; Yuill & Oakhill, 1991). When word reading and reading comprehension difficulties occur together, problems with understanding can arise, because labored decoding of words leaves the reader with insufficient processing capacity to compute the relations between successive words, phrases, and sentences to construct a coherent and meaningful representation of the text (e.g., Perfetti, 1985).

However, good word reading does not necessarily ensure good comprehension. Approximately 10% of (British) schoolchildren have adequate word reading skills but poor reading comprehension (Stothard & Hulme, 1996; Yuill & Oakhill, 1991). Furthermore, these children demonstrate a general deficit in text comprehension, performing poorly on assessments of both reading and listening comprehension (Cain, Oakhill, & Bryant, 2000; Stothard & Hulme, 1992). Thus, word reading deficits are not an obvious source of their reading comprehension difficulties. Furthermore, Oakhill, Cain, and Bryant (2003a) recently demonstrated that entirely different skills underlie competence in word reading and competence in comprehension. Longitudinal research by de Jong and van der Leij (2002) yielded a similar conclusion: Early phonological skills were important predictors of later word reading skill, whereas word reading ability, vocabulary knowledge, and listening comprehension explained growth in reading comprehension.

In this chapter, we address two main questions. First, are the different skills and abilities that are known to be related to comprehension skill (e.g., inference-making, comprehension monitoring, and understanding of text structure) separable constructs, or do they derive from some more general construct that underlies comprehension skill? We present data from our longitudinal study that enable us to determine the relative independence and interrelations between the different subskills of comprehension. Second, is there any evidence that the skills that are correlated with comprehension skill might be causally related to it? To address this question, we present converging evidence from studies that have used three (complementary) ways in which causality can be explored: (a) by comparison of poor comprehenders with a younger comprehension–age match group, (b) by longitudinal studies, and (c) by training studies. We need to have evidence of which skills and abilities have a causal influence on the development of reading comprehension if we are to make progress in the remediation of reading comprehension problems and development of effective comprehension strategy instruction.

COMPONENT SKILLS OF COMPREHENSION

Many skills may contribute to overall comprehension performance. Some accounts of comprehension focus on those skills that contribute to the meaning-construction aspects of the task. For example, Palincsar and Brown (1984) identified six different component skills that make up comprehension ability, including the activation of relevant background knowledge, generation of inferences, and monitoring of both ongoing comprehension and the internal consistency of the text. Other researchers, such as Carr, Brown, Vavrus, and Evans (1990) and Perfetti (e.g., Perfetti, 1985, 1994; Perfetti, Marron, & Foltz, 1996) have included both lower level skills, associated with efficient word reading, and higher level skills and processes, associated with meaning construction, in their accounts of the factors that contribute to reading comprehension ability.

Within the information-processing framework, slow or inaccurate word reading is proposed to affect comprehension by using up too much processing capacity, leaving little remaining for text comprehension processes such as integration (Perfetti, 1985). In accordance with this theory, word reading is the best predictor of reading comprehension level in the early years (e.g., Juel et al., 1986), but other skills become the more important predictors of comprehension level as word reading ability develops through experience (Curtis, 1980; Saarnio, Oka, & Paris, 1990). There are three higher level skills associated with meaning construction that we focus on in this chapter: (a) inference-making, (b) comprehension monitoring, and (c) understanding text structure.

Inference-Making

Inference-making is essential to ensure good understanding of a text. The writer does not necessarily state every little detail: That would result in text that was lengthy and possibly boring. Instead, the reader is left to fill in details that are not explicitly stated in the text, either by integrating statements within the text or by incorporating general knowledge with textual information. Developmental studies of inference skills show that young children are able to make the same inferences as older ones but are less likely to do so spontaneously (Barnes, Dennis, & Haefele-Kalvaitis, 1996; Casteel, 1993; Casteel & Simpson, 1991). Furthermore, impaired inference-making skill is related to children's reading comprehension difficulties (Cain & Oakhill, 1999; Cain, Oakhill, Barnes, & Bryant, 2001; Oakhill, 1982, 1984).

Comprehension Monitoring

The ability to monitor one's understanding of a text is an important skill for constructing meaning. If an individual realizes that his or her comprehension is inadequate, then he or she can take appropriate steps to remedy the problem. Comprehension monitoring is often assessed by requiring readers to detect inconsistencies in text, such as scrambled sentences, contradictory sentences, or statements that conflict with readily available general knowledge. Performance on these tasks improves with age (Baker, 1984; Markman, 1981), and it has been suggested that growth in this skill may be related to children's developing information-processing capabilities (Ruffman, 1996; Vosniadou, Pearson, & Rogers, 1988). Children with reading comprehension difficulties are poor at detecting internal inconsistencies in text (Ehrlich, 1996; Ehrlich, Remond, & Taradieu, 1999). Furthermore, difficulties are more pronounced when the anomalous pieces of information are nonadjacent (e.g., Oakhill, Hartt, & Samols, 2005; Yuill, Oakhill, & Parkin, 1989), indicating that processing capacity may influence the application of this skill.

Understanding Text Structure

Explicit awareness about text structure and the expectations engendered by certain common features of text may be useful aids for readers, helping them to invoke relevant background information and schemas to facilitate their construction of a meaning-based representation. Knowledge about the organization of narrative texts increases throughout middle childhood (e.g., Stein & Glenn, 1982), as does the ability to generate well-structured coherent stories (e.g., Berman & Slobin, 1994). Perfetti (1994) proposed that a possible source of comprehension failure is inadequate knowledge about text structures and genres, which may arise because of insufficient reading experience. Indeed, instruction in expository text structure aids reading comprehension (see chaps. 8 and 14, this volume).

Paris and colleagues (see, e.g., Myers & Paris, 1978; Paris & Jacobs, 1984) have found that knowledge about the purpose of reading and knowledge about the information provided by conventional features of text are related to both age and reading comprehension. In particular, they found that older readers and better comprehenders were better able to explain the sorts of information that may be provided by the introduction and ending of a text. Children with specific comprehension difficulties demonstrate impairments in their ability to structure stories (Cain & Oakhill, 1996; Cain, 2003) and have impoverished knowledge about the information contained in certain features of text, such as titles (Cain, 1996).

Lower Level Skills

Other skills, such as vocabulary and knowledge of word meanings, might also limit comprehension, for obvious reasons. Word knowledge is highly correlated with reading comprehension ability in both children and adults (Carroll, 1993), but the relation between the two is not clear. Limited vocabulary knowledge does not always impair comprehension (Freebody & Anderson, 1983), and vocabulary knowledge per se does not appear to be sufficient to ensure adequate comprehension of larger units of text (e.g., Pany, Jenkins, & Schreck, 1982). The relation between vocabulary and reading comprehension seems to change with age: Torgesen, Wagner, Rashotte, Burgess and Hecht (1997), for example, found that second-grade vocabulary explained 24% of the variance in fourth-grade reading comprehension, whereas third-grade vocabulary explained 43% of the variance in fifth-grade reading comprehension. This finding illustrates the point that the precise relation between various skills and abilities and reading comprehension may change developmentally (see also chap. 2, this volume).

Several researchers have demonstrated that children can experience text comprehension difficulties even when vocabulary knowledge is controlled for (e.g., Cain, Oakhill, & Lemmon, 2005; Ehrlich & Remond, 1997; Stothard & Hulme, 1992; see Cain & Oakhill, 2004, for a review). Thus, reading comprehension clearly involves more than simply accessing the meanings of the individual words in the text.

Although simple vocabulary knowledge may not be a strong determinant of comprehension skill, individuals who possess a rich and interconnected knowledge base may comprehend text better than those whose representations are sparse. Thus, if word meanings are poorly represented in semantic memory, less information will be accessed, and perhaps fewer relations between concepts will be made, than if a rich semantic representation for word meaning exists. Nation and Snowling (1998) reported differences between good and poor comprehenders on a measure of semantic fluency: Poor comprehenders produced fewer instances than did good comprehenders. In a further study, they compared the priming of words that were related by category but differed with respect to associative strength occurrence in both good and poor comprehenders (Nation & Snowling, 1999). Good comprehenders showed priming for both types of word pairs, whereas poor comprehenders showed priming only for word pairs that had high associative strength as well as a category relation.

At first glance, these studies suggest that children with specific comprehension problems might also experience semantic weaknesses. However, it is unclear whether this is indeed the case, because in Nation and Snowling's (1998, 1999) studies the good and poor comprehenders were not matched for word reading accuracy or for vocabulary skills. In fact, vocabulary differences discriminated between the groups as well as did comprehension skill differences: In a second analysis, in which high and low vocabulary groups were compared, the results were parallel to those found when comprehension skill groups were compared. Thus, it is possible that the good comprehenders' superior performance on this task was due to their better word reading and/or vocabulary skills rather than their superior discourse-level comprehension.

Indeed, Cain, Oakhill, and Lemmon (2004) found that when good and poor comprehenders are matched for both word reading accuracy and sight vocabulary, they produced comparable numbers of exemplars on a semantic fluency test. Unfortunately, there is as yet little research that has explored the causal links among vocabulary, word-level knowledge, and comprehension skill (although our longitudinal study, which we discuss shortly, addressed the relation between single-word reading and vocabulary and comprehension skill).

Children With Specific Reading Comprehension Difficulties

The focus of our own work has been on those skills and abilities that do (and do not) differentiate good from poor comprehenders who do not have a problem in reading single words. Much of our work has focused on three main areas in which poor comprehenders have difficulties (when compared to same-age good comprehenders, who have the same level of word reading skill): (a) inference-making, (b) comprehension monitoring, and (c) understanding of story structure. However, it is of course not sufficient to have a list of skills and abilities that differ between good and poor comprehenders. If children's comprehension is to be fostered developmentally, and poor comprehenders are to be helped, then researchers need to know which of the many associates of comprehension skill are likely to be causally implicated in comprehension development. The fact that good and poor comprehenders differ in, say, inference-making skill tells us nothing about the direction of the link. It may be that skill in inference-making is a fundamental ability and that being good at it fosters comprehension development (i.e., inference skill causes comprehension development). However, it could equally well be the case that being good at text comprehension, and having lots of experience of reading for comprehension, fosters inference-making skills (i.e., comprehension causes inference skill development). Or, of course, it is possible that there is reciprocity in skill development across time. In the next section, we outline the main methods that have been used to explore causality in reading research.

METHODS FOR EXPLORING CAUSALITY

A problem for researchers of reading development is that there is no one best method to establish which variables are causal. Each method has its limitations, so it is prudent to use all methods and to look for converging evidence as to which particular variables are causally implicated in comprehension skill. The three methods that have been used, each of which we discuss and illustrate in more detail, are (a) the comprehension–age match (CAM) design, (b) longitudinal studies, and (c) training studies.

TABLE 3.1
Characteristics of Typical Groups of Participants in the Comprehension-Age Match Design

	Less-skilled	*Skilled Control*	*CAM control*
Age (years; months)	7;8	7;8	6;8
Neale word reading Accuracy (y;m)*	7;11	7;10	6;8
Neale Comprehension (y;m)	6;7	8;2	6;8

**Note.* The Neale Analysis is a standardized test of reading, which provides measures of both word reading accuracy and reading comprehension.

The Comprehension–Age Match Design

The CAM design is analogous in its logic to the reading–age match design pioneered by Bryant and colleagues (see Bryant & Goswami, 1986, for an overview). The reading–age match design has proved a popular method for identifying candidate causes of backward readers' delayed reading and, for similar reasons, the CAM design has been developed to study causes of delayed comprehension (Cain et al., 2000; Stothard & Hulme, 1992). The CAM design requires three groups of participants: (a) good and (b) poor comprehenders, who are matched in age and single-word reading ability, and (c) an additional group of younger, normally developing comprehenders. This younger group is selected so that their comprehension skill is at the same absolute level as that of the older poor comprehenders but is normal for their age. Thus, a group of poor comprehenders, aged 9, with comprehension skill comparable to that of 7½-year-olds, would necessitate the use of a comparison group of normally developing 7½-year-olds. An example of the three groups of participants used in such studies is provided in Table 3.1.

The logic of the design is as follows: If the younger CAM group performs better on some task known to differentiate between good and poor comprehenders (e.g., ability to draw inferences from text), this difference cannot be a product of the two groups' comprehension levels, because the CAM group and the poor comprehenders are matched on absolute level of comprehension skill. It is therefore more likely that the difference is associated with the cause of the poor comprehenders' delay. This is, of course, a strong test, because the CAM group is necessarily composed of younger children who have poorer word reading skills than the poor comprehenders.

The CAM design can provide relatively quick and inexpensive evidence, but it can be used only to rule out a causal link in one direction (from comprehension ability to the skill in question) and cannot be used to prove a link in the opposite direction. Thus, other, more expensive and time-consuming methods, such as longitudinal and training studies, are needed to complement this approach. We present the findings from our own studies using this design in the next section.

Longitudinal Studies

Longitudinal studies that track the course of changes in comprehension skill can provide additional information about the causal relations among the components of comprehension and thus about the course of development. The aim of longitudinal studies is to measure sets of skills and abilities at different time points and then to assess (using multiple regression or causal modeling, e.g., structural equation modeling) whether some variables are better predictors of later comprehension skill (and growth in comprehension skill with time) than are others and that particular (early) variables are better predictors of comprehension skill across time than (early) comprehension skill is a predictor of that variable. In addition, de Jong and van der Leij (2002) argued that if a certain variable can be shown to predict reading across time over and above the autoregressive effect of reading itself, then such a pattern provides evidence for a causal link between that variable and comprehension skill.

Longitudinal studies have their limitations, like any other research design. One restriction of this design, particularly when it is used to study growth in academic skills such as reading, is that it is not always possible to measure the skill of interest at the earliest time point in order to take its autoregressive effect into account in the statistical analysis. For example, most 5-year-olds would score poorly on a measure of reading comprehension because their word reading skills would limit their performance. However, it is possible to take measures of the underlying construct, for example, narrative comprehension itself. For an example of this approach, see chapter 2 in this volume.

Another limitation is that a skill identified as a predictor of later competence might achieve its effects indirectly, through a mediating skill. For example, early phonological awareness might appear to predict later reading comprehension, but the relationship might be only indirect because phonological awareness is a strong predictor of a more likely causal influence, word reading ability. We took account of such possibilities in the longitudinal study that we discuss later. In this study, Oakhill, Cain, and Bryant (2003b) assessed the reading comprehension skills of children when they were aged 7–8, 8–9, and 10–11. At each time point, they also assessed basic cognitive skills that are known to be associated with reading comprehension so that they could control for the influence of these factors in the analysis. These control measures were verbal and performance IQ (7–8 and 10–11 years only), vocabulary knowledge (receptive and written), and word reading skill. Many other skills were assessed in this work to investigate other aspects of language development, but we focus on only those that are central to the findings discussed here. Finally, they included measures of the three comprehension-related skills discussed earlier: (a) inference-making, (b) comprehension monitoring, and (c) understanding of text structure.

Training Studies

The logic of training studies is, at first glance, straightforward: If a group trained on some skill this is a candidate cause of good comprehension improves on a standardized measure of comprehension more than a control group not trained on that specific skill (but given some other training of a comparable duration), then it is assumed that the trained skill is causally implicated in the improvement in comprehension. For example, if training poor comprehenders to make inferences not only improved their inference making ability but also led to greater gains in their overall reading comprehension skill compared with a group of good comprehenders, one might assume that inference-making deficits were the cause of the poor comprehenders' reading comprehension problems.

However, some caution must be exercised in the interpretation of such results, as it is not always the case that a training effect occurs (directly) because of an improvement in the skill being trained. Most studies do not test this assumption. For example, if instruction is given in questioning, and free recall is subsequently found to improve, is this a direct result of an improvement in questioning, or is the improvement mediated by an improvement in, say, semantic activation (or some other skill or ability)? Unless researchers are clear which sorts of training are having a direct rather than a mediated effect, they will not be able to determine the most efficacious training methods (e.g., in the preceding example, it might be more efficient to train children in rapid activation of semantic information if that were shown to be a mediating variable).

The effects of possible mediating variables were explored in a recent study by Guthrie et al. (2006), and we use this study as an illustration of how such effects can be investigated. Previous studies by Guthrie and colleagues have shown that Concept-Oriented Reading Instruction (CORI) increases reading comprehension compared with strategy instruction or traditional instruction (Guthrie et al., 2004). However, these studies did not determine whether the positive influence of CORI on reading comprehension arose because of the students' increased level of engagement during instruction. The positive effects may have arisen because students learned more strategies, such as questioning, summarizing, and searching text. Or, it could be that CORI increased comprehension by its effects on oral reading fluency. Guthrie et al. (2006) again showed that CORI had a positive effect on reading comprehension, relative to other training conditions, and that training group also influenced reader engagement, with higher levels of engagement in the CORI group. However, in a further analysis, in which they included the measure of reader engagement (teachers' ratings) as a covariate, the significant effect of training condition was no longer present. Thus, the positive effects of CORI on subsequent reading comprehension were entirely mediated by the students' level of engagement during instruction. The finding that students' engagement mediates the effects of instructional practices raises reservations about the view that reading instruction is effective because of its effects on particular cognitive strategies (see, e.g., National Reading Panel, 2000) and, more generally, highlights the fact that one cannot necessarily conclude that instruction in a particular cognitive skill is effective in improving reading comprehension (if it does) because of the improvement in the cognitive skill.

EVIDENCE FOR CAUSAL RELATIONS

In this section, we draw on our own research that has investigated causal relations between inference-making, comprehension monitoring, understanding text structure and reading comprehension. We evaluate the evidence for causal relations from studies that have used the three designs that can address causality: comprehension–age match comparison, longitudinal and training studies.

Inference-Making

We have investigated causal relations between reading comprehension and inference-making using all three research designs. We have previously shown that good and poor comprehenders differ in their propensity to make two different types of inferences (Oakhill, 1984). These are so-called *text connecting inferences*—which are required to integrate information from different parts of the text, to establish local coherence, and *gap-filling inferences*, in which information from outside the text (general knowledge) is incorporated with information in the text to fill in gaps in missing details and to help formulate a coherent representation of the text as a whole.

We explored children's ability to answer these types of question in a study that used the CAM design (Cain & Oakhill, 1999). In all of the studies using the CAM design reported in this chapter, the good and poor comprehenders were aged 7 to 8 years, and the children in the CAM group were aged 6 to 7 years. Children read short stories, each of which was followed by four inference

TABLE 3.2
Example Story and Questions From the Comprehension-Age Match Study of Inferences

Debbie was going out for the afternoon with her friend Michael. By the time they got there they were very thirsty. Michael got some drink out of his duffel bag and they shared that. The orange juice was very refreshing. Debbie put on her swimming costume but the water was too cold to paddle in, so they made sandcastles instead. They played all afternoon and didn't notice how late it was. Then Debbie spotted the clock on the pier. If she was late for dinner her parents would be angry. They quickly packed up their things. Debbie changed and wrapped her swimming costume in her towel. She put the bundle in her rucksack. Then they set off for home, pedaling as fast as they could. Debbie was very tired when she got home, but she was just in time for dinner.

Questions requiring retrieval of literal information:
Who did Debbie spend the afternoon with?
Where was the clock?

Questions requiring text-connecting (bridging) inferences:
Where did Michael get the orange juice from?
Where did Debbie put her towel when she packed up her things?

Questions requiring gap-filling inferences:
Where did Debbie and Michael spend the afternoon?
How did Debbie and Michael travel home?

questions (two of each of the types described earlier) and two literal questions. An example of a story, with questions, is shown in Table 3.2. The groups did not differ in their ability to answer the questions tapping literal information in the text. However, the poor comprehenders were poorer at answering both types of inferential questions than good comprehenders, and they were worse than the CAM group at text-connecting inferences. The differences between the poor comprehenders and the CAM group cannot be the result of any differences in comprehension level, because the two groups were matched for absolute reading comprehension level. This result strongly suggests that skill at drawing inferences does not simply arise as a by-product of reading comprehension skill; instead, it suggests that impaired inference-making skill might be causally related to poor reading comprehension.

In our longitudinal study, we explored the causal relations between reading comprehension skill and inference-making abilities further. We used stories similar to those used in the CAM study to assess children's inference-making skills when they were aged 8 to 9 years and 10 to 11 years. At the first time point, when children were aged 7 to 8, we used a different task, which assessed children's ability to integrate different sentences within short texts (Oakhill, 1982). Inference-making was strongly related to reading comprehension at each time point over and above our basic control measures of word reading, vocabulary, and IQ. The inference and integration measure taken at Time 1 was significantly correlated with later reading comprehension ability, but it did not explain reading comprehension once we had taken IQ, word reading, and vocabulary knowledge into account. However, the ability to generate inferences when aged 8 to 9 was significantly related to reading comprehension 2 years later, over and above our general ability control measures. Furthermore, inference making at age 8 to 9 explained later reading comprehension even after we had taken earlier reading comprehension ability into account. These data strongly suggest that early inference skills are causally related to the development of reading comprehension.

Two training studies support this proposed direction of causality. In the first, Yuill and Joscelyne (1988) taught 7- to 8-year-old children to make inferences from key "clue" words in deliberately obscure texts. In one story, the text does not state explicitly that the main character was lying in the bath, but this setting can be inferred from words such as *soap*, *towel*, and *steamy*. After training, the children were tested on similar types of stories. The poor comprehenders benefited more from the training than the skilled group. In another study, Yuill and Oakhill (1988) trained children of the same age to make the same type of lexical inference and to generate questions to test their understanding. This study combined training in inference-making and comprehension monitoring. The poor comprehenders who had received this training made substantial gains in comprehension on a standardized measure of reading comprehension, relative to the good comprehenders.

Comprehension Monitoring

Work investigating the causal relations between reading comprehension and comprehension monitoring is scarce. We know of no work that has used the CAM design. We have successfully trained monitoring skills, by teaching children how to form questions to check their understanding and identify the need to generate inferences, in the study described earlier (Yuill & Oakhill, 1988). However, our attempts to teach comprehension monitoring skills alone, rather than in conjunction with inference-making skills, have been less successful (Yuill & Oakhill, 1991). One reason for this may be that longer periods of training are required, or that monitoring is inextricably linked to comprehension skill, such that it cannot be improved unless comprehension itself is improved by other means.

The ability to monitor comprehension was assessed in our longitudinal study using an inconsistency detection task. At each time point, children were presented with short stories, some of which contained inconsistent lines.

TABLE 3.3
Example of Text Used in Inconsistency Detection Task

Moles

Moles are small brown animals and they live underground using networks of tunnels.
Moles cannot see very well, but their hearing and sense of smell are good.
They sleep in underground nests lined with grass, leaves and twigs.
Moles use their front feet for digging and their short fur allows them to move along their tunnels either forwards or backwards.
They mainly eat worms but they also eat insects and snails.
Moles are easily able to find food for their young because their eyesight is so good.

_____This passage makes sense, it does not need to be changed.
_____This passage does not make sense, it needs to be changed

An example is given in Table 3.3. The task was to underline the lines that "did not make sense." The length of the stories was increased at each time point to ensure that the task had a suitable level of difficulty as the children got older.

Performance on this task was related to concurrent reading comprehension levels at each time point. This relation was evident even after the influence of word reading, vocabulary, and IQ had been taken into account. One possibility that we considered was that processing capacity—that is, working memory capacity—might influence performance. It was particularly important to consider this possibility because the lines of inconsistent text were not adjacent, so the inconsistent formation had to be compared over several lines of text (see the example in Table 3.3). We assessed working memory capacity using a sentence-based and a numerical task that both required simultaneous processing and storage of information. Performance on the comprehension monitoring task explained variance in reading comprehension skill even after we had controlled for working memory capacity, suggesting that processing capacity in itself was not sufficient to explain the relation between the two skills (Cain, Oakhill, & Bryant, 2004).

Similar to our findings for inference-making performance, we found that early comprehension monitoring skill was related to later reading comprehension ability and that comprehension monitoring at 8 to 9 years predicted later reading comprehension skill over and above the contribution of other basic level skills, namely, word reading, IQ, and vocabulary, and even after earlier reading comprehension skill had been taken into account. Again, we have evidence that a skill that is specifically related to the meaning-based aspects of text processing might be causally related to the development of reading comprehension.

TABLE 3.4
Examples of Story Productions From the CAM Production Study

Non story ('the holiday'):

Once upon a time there was a girl and she went on holiday.

Intermediate story ('the seaside'):

It was a lovely day. The family decided to go down to the seaside. They saw lots of people there. The baby was making a sandcastle. The older children were playing in the sea. The mum and dad had their last swim before they went home.

Complete story ('the circus')

One day there was a girl named Gigi. She went to a circus. She saw loads of things she liked. The best thing she liked was the lion... In the morning she found the lion she liked the best, what was from the circus, walking up and down the road by her house. So she said, "Mum, the lion I saw in the circus last night is walking up and down in the road". So her mum said, "Go and get some kippers and try and head him back to the circus and put him in the cage". So she did. The end.

Understanding Text Structure

In our work on text structure, we have focused on children's understanding of story structure, because they have more experience of this genre in early childhood. We have used the CAM design to investigate the relations between reading comprehension and the ability to structure stories in two studies. In the first, children's ability to produce well-constructed stories was compared in two different conditions: either (a) with a "topic prompt," such as "Pirates," or (b) with a series of pictures that essentially "told a story" (Cain & Oakhill, 1996). The study had various aims, but the present interest is in differences in the quality of the story structures between the groups. The children were required to produce short stories verbally, their productions were transcribed, and the productions were classified using a three-way classification that reflected the level of interconnectedness and causality. The categories were nonstories, intermediate, and well-formed stories. The children's productions were given a score according to which category they fell into, and these scores were used as the dependent variable in the analysis. Some examples of stories from the different categories are shown in Table 3.4.

The poor comprehenders produced less well-structured stories than did good comprehenders, and they were worse than the CAM group in a condition where only a topic for the story was provided. Looking at the examples provided, one might wonder if this pattern of results was confounded with the amount that children produced (in the examples, there is a correlation between length and quality of the stories). However, this did not seem to be the case in general: In the topic prompt condition (the one that discriminated between groups), the good comprehenders produced the shortest stories (on average, 119 words, compared with 147 for the CAM group and 172 for the poor comprehenders). In a second study, designed to look at the support offered by the prompts in more detail, we replicated the main finding that the CAM group produced stories with more integrated structures than the poor comprehenders when provided with only a topic prompt (Cain, 2003). Together, the results of these CAM design studies strongly suggest that the ability to produce causally coherent stories is not simply a by-product of reading comprehension skill, because the poor comprehenders were worse on this task relative to children at the same absolute level of reading comprehension skill (the CAM group). These findings have identified impaired story structure skills as a likely cause of poor comprehension.

It was not possible to include a story production task in our longitudinal study, because of the time required to record, transcribe, and score the individual stories. Instead, we adapted a task used by Stein and Glenn (1982) to measure children's ability to structure stories. In our version, children were given three short stories that had been cut up into their constituent sentences. The sentences were presented on individual strips of paper in a randomized order, and the task was to put the sentences in the correct order so that the story made sense. The number of sentences in each story was increased as the children got older. We calculated concordance ratings, which reflect the degree to which the arranged sequence matched the correct sequence of sentences. We also took measures of children's knowledge about the sorts of information contained in story titles and the beginnings and ends of stories.

Performance on the story anagram task was significantly correlated with reading comprehension at each time point, but it explained only unique variance (over and above our basic control variables) when children were aged 10 to 11 years. Knowledge about the information provided by titles, beginnings, and endings was more strongly related to concurrent measures of reading comprehension skill. When we looked at how these variables predict later competence in reading comprehension, we found strong indications that knowledge about text structure is an important factor in reading comprehension development. Knowledge about the information contained in story titles at 7 to 8 years predicted reading comprehension level 1 year later, over and above our basic control variables and reading comprehension assessed at Time 1. Similarly, the ability to rearrange the story sentences into a coherent order at Time 1 predicted later reading comprehension at Times 2 and 3, over and above the other variables. Again, we find evidence that a skill related to the construction of a text's meaning is related to the development of reading comprehension skill.

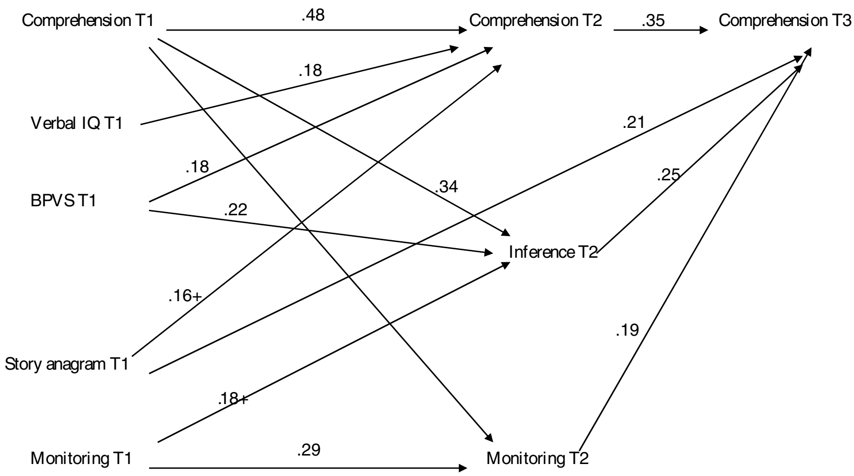


Figure 3.1. Path analysis based on data from the longitudinal study by Oakhill, Cain, and Bryant (2003b). Variables measured at Time 1 (T1; age 7–8) predict variables at Time 2 (T2; age 9–10) and Time 3 (T3; age 11–12). The paths shown indicate predictors that were significant after the effects of all other variables were removed: a reading comprehension measure (comprehension), a picture vocabulary test (BPVS; Dunn, Dunn, Whetton & Burley, 1998), Verbal IQ (Wechsler Intelligence Scale for Children—Revised; Wechsler, 1974), detection of contradictions in text (monitoring), a sensitivity to story structure measure (story anagram), and an inference measure (inference). Paths that linked T1 and T2 variables, but that did not link to T3 comprehension (directly or indirectly), have been excluded for clarity. The analyses were based on standardized data; thus, the path coefficients shown are directly comparable. Marginally significant paths are marked with a plus symbol (+).

An Integrated Model of Comprehension Development

Thus far, we have presented findings from our longitudinal study related to the three specific skills that have been the focus of this chapter: (a) inference-making, (b) comprehension monitoring, and (c) knowledge of text structure. Now we present a causal path diagram to show the pattern and strength of relations among all of the skills across time. This type of analysis identifies whether the three specific skills have separable influences on reading comprehension and reading comprehension development. Furthermore, it demonstrates whether the links between inference-making, comprehension monitoring, and knowledge about text structure and reading comprehension are mediated by the control variables that we considered, namely, word reading, IQ, and vocabulary. The final causal path diagram, with only significant paths included, is shown in Figure 3.1.

Initial comprehension skill was a strong predictor of later comprehension, and verbal ability (vocabulary and verbal IQ) also made significant contributions to the prediction of comprehension ability across time. Nevertheless, three distinct predictors of comprehension skill emerged, either through direct or indirect links: (a) answering inferential questions, (b) monitoring comprehension, and (c) understanding story structure. These factors predicted comprehension at a later time even after the autoregressive effect of comprehension (the prediction of comprehension at later times from comprehension at earlier times) was controlled. In addition, it was not the case that one skill (e.g., comprehension monitoring) mediated the links between the other skills (e.g., inference-making and knowledge of text structure and reading comprehension). With word reading as the dependent variable, the pattern was quite different. The significant predictors were previous measures of word reading and a measure of phonological skills taken at Time 1. From these analyses, a picture of skill development emerges in which certain components of comprehension are predictive of general comprehension skill. Earlier abilities in inference skill, comprehension monitoring, and knowledge of text structure all predict a later global assessment of comprehension skill (but not word reading) over and above the autoregressive effect of comprehension skill and in competition with measures of general ability (Verbal IQ, vocabulary, and working memory).

This analysis confirms that a set of higher level comprehension components, which, on theoretical grounds, ought to be instrumental in the growth of reading comprehension skill, may indeed be instrumental. These findings suggest that comprehension does not necessarily develop automatically once word reading is proficient but that it is dependent on different skills and may need specific teaching. In relation to this final point, our findings sit very well with Kendeou et al.'s findings (chap. 2, this volume) that reading comprehension is specifically predicted by earlier language comprehension ability.

CONCLUSIONS AND IMPLICATIONS FOR STRATEGY TRAINING

Our review of the literature and our recent research has demonstrated that reading comprehension is a complex task that draws on a range of skills and processes. It is not surprising, therefore, that children with reading comprehension difficulties show impairments on a range of language tasks. As a result, it might appear at first to be an impossible task to either identify which skill(s) should be targeted in strategy instruction or to develop instruction that can target such a wide variety of skills.

In the studies we have described, we have tried to identify at least some of the skills and abilities that might be causally implicated in comprehension ability and which, therefore, might be fruitfully trained. The three abilities we have identified, and shown to be consistently and reliably related to comprehension skill, are (a) inference-making, (b) comprehension monitoring, and (c) understanding story structure. In addition, we have provided evidence from the CAM, longitudinal, and training studies showing that each of these abilities may be causally implicated in comprehension development. Although we do not always have convergent evidence from all three research designs for each skill, the data so far strongly suggest that early competence in each of the three areas of ability helps reading comprehension to develop.

However, the leap from these findings to the type of training that is likely to be most effective is not straightforward. There is some evidence that these skills are independent: For example, in our longitudinal study there was no evidence that one specific skill mediated the links between the other skills. However, even if the skills are at least to some extent independent, it is not clear whether they should be trained separately or in tandem and whether it is possible to train skills separately. Is it possible, for instance, to train awareness of the need to generate inferences, without emphasizing the need to monitor comprehension? How can we teach children to represent the causal relations between events in a narrative without including some instruction in inference-making and monitoring skills?

In addition, we need to consider the essence of successful comprehension and how these distinct skills contribute to it. Successful comprehension involves the construction of a clear, complete, and integrated representation of a text's meaning, a *mental model* of the text (Johnson-Laird, 1983). It is clear that successful comprehension can be achieved only if that is the reader's goal. We have speculated previously that good and poor comprehenders may have different goals for reading. Poor comprehenders are capable of generating inferences; they simply do not generate sufficient inferences to ensure adequate comprehension of text (Cain & Oakhill, 1999). Relative to good comprehenders, poor comprehenders are less likely to adapt their style of reading to meet different task aims (Cain, 1999), and they tend to view reading as a word decoding activity rather than one of meaning-making (Cain, 1999; Yuill & Oakhill, 1991). Thus, one aim of remediation may be to teach poor comprehenders about the goals of reading and the aim to strive for a referentially and causally coherent representation of the text's meaning.

This suggestion maps onto to the idea of a reader's standard for coherence—caring that the text makes sense (e.g., Perfetti, Landi, & Oakhill, 2005; van den Broek, Risden, & Husebye-Hartman, 1995). A relation between inference-making, comprehension monitoring, story production and reading comprehension could be, at least in part, mediated by a reader's standard for coherence. When the goal is to derive a coherent representation of meaning, inferences are made to keep the text coherent. When coherence is a goal, inconsistencies between text elements or between text elements and the reader's knowledge are likely to be resolved rather than ignored or not noticed. When coherence is a goal, the point and structure of the text as a whole will be appreciated by the reader and will, in turn, guide and reinforce the reader's comprehension.

Better readers maintain higher standards for coherence (van den Broek & Kremer, 1999); thus, an obvious suggestion is that remediation programs should teach poorer readers to improve their coherence standards. There are at least three issues with this argument. First we do not know how the absence of a "drive for coherence" encapsulates the comprehension problems that have been documented. Although it seems plausible that many of the comprehension difficulties that are observed (problems with inference-making, monitoring comprehension, and understanding story structures) might derive from an inadequate drive for coherence, it is an empirical issue whether this more overarching construct can encompass the problems that poor comprehenders have with these various skills. One source of evidence that would help to address this question would be the existence of children who have problems in some, but not all of, the components that would be expected to be associated with the lack of a drive for coherence: for example, children who are good at comprehension monitoring in the presence of poor inference skills.

Second, there are empirical issues to be addressed as to how to go about encouraging children to make coherence links. Paris and colleagues have developed an effective program to raise children's awareness about the purpose of reading, but it has met with limited success in improving standards of comprehension (e.g., Paris & Jacobs, 1984). Thus, it might be more beneficial to teach them some of the component processes that contribute to the construction of a coherent text.

Third (and related to the last point), because comprehension monitoring is both a product and an essential component of the drive for coherence, and because effective monitoring will depend on the reader having a high coherence standard, poorer readers may not benefit from training in strategies designed to improve their drive for coherence, because they may not know when to apply them (because they are poor at comprehension monitoring).

There is mounting evidence that certain methods are effective in improving comprehension skill. Rather than reviewing specific programs, we note the wide extent of such research, drawing on a comprehensive review of research on reading (National Reading Panel, 2000). The National Reading Panel report (2000) identified seven categories of comprehension instruction that have solid evidence for their effectiveness. These seven include procedures that we characterize as drawing the reader into a deeper engagement with the text, that is, active processing. They include comprehension monitoring, question answering (teacher-directed questions) and question generation (student self-questioning), the use of semantic organizers (students making graphic representations of text), and student summarization of texts. Instruction in understanding story structures was also judged to be effective.

The procedures that the National Reading Panel report suggests are effective are consistent with the comprehensions skills we have reviewed in this chapter. Active engagement with the meaning of text helps the reader to represent the text content in a way that fosters both learning (as opposed to superficial and incomplete understanding) and an attraction to reading. All these are strategies that encourage readers to adopt a high standard for coherence. We also note that one feature that many successful instructional programs have in common is that they train and encourage students to recognize the important connections in a text (e.g., Reciprocal Teaching, Palincsar & Brown, 1984; and Reading Recovery, Clay, 1991). Thus, our conclusions about which skills might be causally implicated are consistent with the conclusions of the National Reading Panel report, and with other observations, on what works to improve comprehension.

FINAL THOUGHTS

We begin this section with a reminder of some important methodological issues that we have discussed earlier. One is the importance of trying to establish causality and, in particular, direct causality, to maximize effective training. As we suggested earlier, it is not sufficient to have a list of skills and abilities that correlate with reading comprehension level or that are impaired in poor comprehenders. If we wish to foster the development of children's comprehension and remediate poor comprehenders, we need to establish which of the many skills associated with reading comprehension are likely to be causally implicated in comprehension development.

A second related methodological issue is to be aware of the limitations of the methods to establish causality. We outlined converging methods for investigating causal links. The CAM design is the most limited in scope: It may be used to identify potential causal correlates but cannot in itself establish causality. Causal conclusions can be drawn from longitudinal and training studies, provided appropriate controls have been included. With longitudinal studies, it is important to show that the variable of interest predicts reading across time over and above the autoregressive effect of reading itself. With training studies, assumptions should not be made about the efficacy of training a particular skill without a thorough investigation of possible mediators between that skill and the outcome variable.

We end with some questions for which we do not have definitive answers but which we think would be useful for those who conduct training studies to consider. First, can comprehension best be trained by teaching individual comprehension skills, or is a more integrated approach better? The National Reading Panel is agnostic on this point, stating simply that improvement of scores on standardized comprehension tests may require training multiple strategies in combination. In our own work described earlier, we found that teaching children to make inferences in conjunction with comprehension monitoring skills was more successful that training monitoring skills alone (Yuill & Oakhill 1988, 1991). Clearly, more data are needed that address this training issue.

Second, do some individuals benefit more from training in some skills rather than others? For instance, in a recent comparison of good and poor comprehenders, we found some interesting differences in a group of children with poor comprehension (Cain & Oakhill, in press). The groups of 7- to 8-year-old good and poor comprehenders were compared on a range of comprehension fostering and text processing skills, such as inference-making. Comparisons between the groups confirmed previous findings that poor comprehenders in general experience weaknesses on these tasks. However, we also found that some poor comprehenders had average or even good performance on these tasks relative to peers. Our results concur with those obtained by Cornoldi, De Beni, and Pazzoglia (1996), in that no clear fundamental weakness was apparent. Instead, this work indicates that group comparisons may obscure crucial weaknesses in the individual.

Where does that leave us with respect to strategy instruction? Should we tailor intervention programs to the specific weaknesses presented by each child, or should we develop interventions that include work on a number of comprehension-related skills? On a positive note, a clear consensus is emerging from different research designs that help to identify the causes of successful comprehension: The skills that predict the development of reading comprehension are distinct from those that predict the development of word reading ability. Thus, we argue that an effective program to remediate comprehension difficulties must include instruction in specific strategies for making inferences, monitoring comprehension, and/or using text structure. Our discussion about standards for coherence and the interrelations between comprehension-fostering skills leads us to suggest that single-skill programs of instruction may not be the most productive way forward. First, it may simply not be possible to teach skills separately. Second, although there is evidence that the different comprehension-related skills are separable constructs and that they are not deficient in all poor comprehenders, each of these skills facilitates the reader's ability to build a coherent representation of meaning. Furthermore, it is not clear how effectively these skills can be taught unless the reader appreciates, or is taught to appreciate, the purpose of reading. A more fruitful approach to intervention may be to develop integrated programs that include instruction in specific strategies embedded within the overarching aim to teach children awareness of the purpose of reading—the goal of comprehension.

REFERENCES

Baker, L. (1984). Spontaneous versus instructed use of multiple standards for evaluating comprehension; effects of age, reading proficiency, and type of standard. *Journal of Experimental Child Psychology*, *38*, 289–311.

Barnes, M. A., Dennis, M., & Haefele-Kalvaitis, J. (1996). The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from six to fifteen years of age. *Journal of Experimental Child Psychology*, *61*, 216–241.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A cross linguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bryant, P., & Goswami, U. (1986). Strengths and weaknesses of the reading level design: A comment on Backman, Mamen and Ferguson. *Psychological Bulletin*, *100*, 101–103.

Cain, K. (1996). Story knowledge and comprehension skill. In C. Cornoldi & J. V. Oakhill (Eds.), *Reading comprehension difficulties: Processes and remediation* (pp. 167–192). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cain, K. (1999). Ways of reading: How knowledge and use of strategies are related to reading comprehension. *British Journal of Developmental Psychology*, *17*, 295–312.

Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Developmental Psychology*, *21*, 335–351.

Cain, K., & Oakhill, J. (1996). The nature of the relationship between comprehension skill and the ability to tell a story. *British Journal of Developmental Psychology*, *14*, 187–201.

Cain, K., & Oakhill, J. V. (1999). Inference making and its relation to comprehension failure. *Reading and Writing*, *11*, 489–503.

Cain, K., & Oakhill, J. V. (2004). Reading comprehension difficulties. In T. Nunes and P. E. Bryant (Eds.), *Handbook of children's literacy* (pp. 313–338). Dordrecht, The Netherlands: Kluwer.

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, *76*, 683–696.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference making ability and their relation to knowledge. *Memory & Cognition*, *29*, 850–859.

Cain, K., Oakhill, J. V., & Bryant, P. E. (2000). Investigating the causes of reading comprehension failure: The comprehension–age match design. *Reading and Writing*, *12*, 31–40.

Cain, K., Oakhill, J., & Bryant, P. E. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*, 31–42.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*, 671–681.

Cain, K., Oakhill, J., & Lemmon, K. (2005). The relation between children's reading comprehension level and their comprehension of idioms. *Journal of Experimental Child Psychology*, *90*, 65–87.

Carr, T. H., Brown, T. L., Vavrus, L. G., & Evans, M. A. (1990). Cognitive skill maps and cognitive skill profiles: Componential analysis of individual differences in children's reading efficiency. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development* (pp. 1–55). New York: Academic.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Casteel, M. A. (1993). Effects of inference necessity and reading goal on children's inferential generation. *Developmental Psychology*, *29*, 346–357.

Casteel, M. A., & Simpson, G. B. (1991). Textual coherence and the development of inferential generation skills. *Journal of Research in Reading*, *14*, 116–129.

Clay, M. M. (1991). *Becoming literate: The construction of inner control*. Portsmouth, NJ: Heinemann.

Cornoldi, C., De Beni, R., & Pazzaglia, F. (1996). Profiles of reading comprehension difficulties: An analysis of single cases. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and interventions* (pp. 113–136). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Curtis, M. E. (1980). Development of components of reading skills. *Journal of Educational Psychology*, *72*, 656–669.

de Jong, P. F., & van der Leij, A., (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading*, *6*, 51–77.

Dunn, L., Dunn, L., Whetton, C., & Burley, J. (1998). *British Picture Vocabulary Scale II*. Windson: NFER-Nelson.

Ehrlich, M. F. (1996). Metacognitive monitoring in the processing of anaphoric devices in skilled and less skilled comprehenders. In C. Cornoldi & J. V. Oakhill (Eds.), *Reading comprehension difficulties: Processes and remediation* (pp. 221–249). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Ehrlich, M. F., & Remond, M. (1997). Skilled and less skilled comprehenders: French children's processing of anaphoric devices in written texts. *British Journal of Developmental Psychology*, *15*, 291–309.

Ehrlich, M. F., Remond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing*, *11*, 29–63.

Freebody, P., & Anderson, R. C. (1983). Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Reading Behavior*, *15*, 19–39.

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., et al. (2004). Increasing reading comprehension and engagement through Concept-Oriented Reading Instruction. *Journal of Educational Psychology*, *96*, 403–423.

Guthrie, J. T., Wigfield, A., Perencevich, K. C., Taboada, A., Humenick, N. M., Lutz, S., & McRae, A. (2006). *A model of the contributions of classroom instruction to students' reading comprehension and reading engagement*. Manuscript submitted for publication.

Johnson-Laird, P. N. (1983). *Mental models*. New York: Cambridge University Press.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*, *78*, 243–255.

Markman, E. M. (1981). Comprehension monitoring. In W. P. Dickson (Ed.), *Children's oral communication skills* (pp. 61–84). London: Academic.

Myers, M., & Paris, S. G. (1978). Children's metacognitive knowledge about reading. *Journal of Educational Psychology*, *70*, 680–690.

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, *39*, 85–101.

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition*, *70*, 81–83.

National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Pub. No. 00-4769). Jessup, MD: National Institute for Literacy.

Oakhill, J. V. (1982). Constructive processes in skilled and less-skilled comprehenders' memory for sentences. *British Journal of Psychology*, *73*, 13–20.

Oakhill, J. V. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, *54*, 31–39.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003a). The dissociation of single-word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes*, *18*, 443–468.

Oakhill, J. V., Cain, K. E., & Bryant, P. E. (2003b, May). *Prediction of comprehension skill in the primary school years.* Paper presented at the biennial meeting of the Society for Research in Child Development, Tampa, FL.

Oakhill, J. V., Hartt, J., & Samols, D. (2005). Comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, *18*, 657–713.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*, 117–175.

Pany, D., Jenkins, J. R., & Schreck, J. (1982). Vocabulary instruction: Effects of word knowledge and reading comprehension. *Learning Disability Quarterly*, *5*, 202–215.

Paris, S. G., & Jacobs, J. E. (1984). The benefits of informed instruction for children's reading awareness and comprehension skills. *Child Development*, *55*, 2083–2093.

Perfetti, C. A. (1985). *Reading ability*. Oxford, England: Oxford University Press.

Perfetti, C. A. (1994). Psycholinguistics and reading ability. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 849–894). San Diego, CA: Academic.

Perfetti, C. A., Landi, N., & Oakhill, J. V. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford, England: Blackwell.

Perfetti, C. A., Marron, M. A., & Foltz, P. W. (1996). Sources of comprehension failure: Theoretical perspectives and case studies. In C. Cornoldi & J. V. Oakhill (Eds.), *Reading comprehension difficulties: Processes and remediation* (pp. 137–165). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Ruffman, T. (1996). Reassessing children's comprehension-monitoring skills. In C. Cornoldi & J. V. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 33–67). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Saarnio, D. A., Oka, E. R., & Paris, S. G. (1990). Developmental predictors of children's reading comprehension. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 57–79). New York: Academic.

Stein, N. L., & Glenn, C. G. (1982). Children's concept of time: The development of a story schema. In W. J. Friedman (Ed.), *The developmental psychology of time* (pp. 255–282). New York: Academic.

Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language comprehension and working memory skills. *Reading and Writing*, *4*, 245–256.

Stothard, S. E., & Hulme, C. (1996). A comparison of phonological skills in children with reading comprehension difficulties and children with word reading difficulties. *Journal of Child Psychology and Child Psychiatry*, *36*, 399–408.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second- to fifth-grade children. *Scientific Studies of Reading*, *1*, 161–185.

van den Broek, P. W., & Kremer, K. E. (1999). The mind in action: What it means to comprehend during reading. In B. Taylor, M. Graves, & P. van den Broek (Eds.), *Reading for meaning* (pp. 1–31). New York: Teachers College Press.

van den Broek, P. W., Risden, K., & Husebye-Hartman, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R. F. Lorch, Jr., & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353–373). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Vosniadou, S., Pearson, P. D., & Rogers, T. (1988). What causes children's failures to detect inconsistencies in text? Representation versus comparison difficulties. *Journal of Educational Psychology*, *80*, 27–39.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children–Revised*. New York: The Psychological Corporation.

Yuill, N., & Joscelyne, T. (1988). Effect of organisational cues and strategies on good and poor comprehenders' story understanding. *Journal of Educational Psychology*, *80*, 152–158.

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge, England: Cambridge University Press.

Yuill, N. M., Oakhill, J. V., & Parkin, A. J. (1989). Working memory, comprehension skill and the resolution of text anomaly. *British Journal of Psychology*, *80*, 351–361.

4

A Knowledge-Based Framework for Unifying Content-Area Reading Comprehension and Reading Comprehension Strategies

Michael R. Vitale
East Carolina University

Nancy R. Romance
Florida Atlantic University

The authors present a knowledge-based perspective for considering reading comprehension as a special case of general comprehension (i.e., cumulative content-area learning). Within this framework, content-area reading comprehension proficiency is approached as a form of expertise for which prior knowledge is a critical element. A multipart knowledge-focused reading comprehension strategy, whose purpose is to amplify the access and organization of prior knowledge with new knowledge for student learners within content-area reading comprehension environments, is presented, and the results of a proof-of-concept study demonstrating the effectiveness of the strategy with Grade 5 students are reviewed. Using the linkage of reading comprehension to general comprehension as a foundation, the authors discuss the use of interdisciplinary approaches to content-area learning as a means for developing reading comprehension proficiency and consider the associated implications for conducting reading comprehension and reading comprehension strategy research that is maximally relevant to advancing an understanding of content-area reading comprehension in applied school settings.

In their recent publication, *Reading for Understanding*, the RAND Reading Study Group (Snow, 2002) reported that the proficiency of students in reading and comprehending subject-matter text has remained a significant educational problem in Grades 4 through 12—the grade levels at which cumulative and meaningful learning in content areas (e.g., science) is emphasized and when reading to learn becomes a critically important student proficiency. A recent National Assessment of Educational progress report (National Center for Educational Statistics, 2000) found that 38% of 4th-graders were unable to read and understand a paragraph from an age-appropriate children's book, a figure that rose as high as 70% in many school districts. Additionally, the RAND report found that international comparisons of performance on reading assessments placed U.S. 11th-graders close to the bottom of all industrialized countries in reading achievement, a finding paralleling that of the Third International Mathematics and Science Study (Schmidt et al., 2002). Even after 20 years of systemic reform initiatives, there is substantial evidence of a continuing achievement gap between low-socioeconomic status, at-risk students who depend on school to learn and their more advantaged peers on both basic skills and content-area achievement (e.g., Florida Department of Education, 2005; National Center for Educational Statistics, 2000; North Carolina Department of Public Instruction, 2005).

As noted in the RAND (Snow, 2002) report and other national reports (e.g., National Reading Panel, 2000), a substantial number of research studies in the fields of reading and educational/instructional psychology have investigated different aspects of teaching reading comprehension (e.g., Block & Pressley, 2002; Farstrup & Samuels, 2002; Gersten, Fuchs, Williams, & Baker, 2001). Yet, in evaluating the state of such research, the RAND report concluded that present knowledge in the field is not sufficient to systemically reform reading comprehension instruction, a finding that suggests serious limitations of such research. Of particular relevance to the arguments advanced in this chapter are two important and interdependent conclusions reported by RAND: (a) that the field of reading has made only minimal progress in the area of content-area reading comprehension and (b) that, although use of reading comprehension strategies can be taught experimentally, the benefits of such strategies have not been found transferable to enhancing content-area reading comprehension in applied school settings.

Using the preceding issues as a context, we present a knowledge-based perspective for considering the enhancement of reading comprehension proficiency through the use of reading comprehension strategies applied in content-area instructional settings by using science as an exemplar of meaningful learning. In doing so, the chapter integrates interdisciplinary research from applied cognitive science that complements and enhances recognized research in the field of reading. Emphasized is how research associated with expertise (Bransford, Brown, & Cocking, 2000) provides a foundation for developing content-area student mastery in a form that can be considered equivalent to meaningful comprehension. Addressing these concerns, we present perspectives in a fashion meaningful to researchers studying cumulative meaningful learning in applied settings and to practitioners working to optimize meaningful student learning. In support of this dual emphasis, the literature cited was chosen to highlight representative aspects of applied cognitive science, reading comprehension, and other related disciplines to which researchers and practitioners alike can relate.

PERSPECTIVES ON READING COMPREHENSION: AN INFORMAL ANALYSIS

The primary argument underlying this chapter is that in content-area learning environments the idea of comprehension is far more general than that of reading comprehension. This distinction is important, because reading comprehension instruction at upper elementary grade levels (Grades 3–5) in the United States has excluded reading materials that require cumulative, meaningful, content-area learning in favor of narrative stories that do not (see Hirsch, 1996; Walsh, 2003). The issue here is not whether reading comprehension proficiency can be engendered as a transferable skill but rather the determination of the conditions through which content-area comprehension can be developed by using reading as a means for learning. An important question yet to be answered through research is the degree to which such cumulative, meaningful, content-area student learning opportunities are necessary for the development of reading comprehension proficiency within and across disciplines.

Consideration of comprehension as more general than reading comprehension magnifies the role of prior knowledge as a primary factor in student meaningful learning. In this regard, the development and subsequent access and use of prior knowledge in reading can be considered in a fashion paralleling their importance in the development of expertise (Bransford et al., 2000) and for cumulative content-area school learning (Hirsch, 1996).

With this perspective in mind, three approaches for linking the study of reading comprehension and content-area learning can be distinguished. Approach 1 emphasizes the development of in-depth content-area understanding as a vehicle for enhancing subsequent learning success (i.e., comprehension) through a variety of instructional activities, of which content-area reading is one. This approach, which may incorporate reading comprehension strategies as part of content-area reading, logically must consider the general enhancement of reading comprehension proficiency as a side effect of meaningful content-area learning. In such an approach represented by content-area learning in science (see Vitale & Romance, 2006a), the emphasis is on having a coherent, concept-oriented curriculum (see Schmidt et al., 2002) within which reading is one of several instructional modes that provides prior knowledge for future learning.

Approach 2, a more traditional approach for the study of reading comprehension (e.g., Block & Pressley, 2002; Farstrup & Samuels, 2002; Gersten et al., 2001), emphasizes the use of narrative curriculum content (i.e., stories) common to basal reading series as a vehicle for developing general student reading comprehension proficiency, typically through the use of reading comprehension strategies. In such an approach, enhancement of student content-area reading comprehension proficiency is a matter of transfer to content-area reading in science and other areas. As noted previously, accomplishing such transfer through this approach has proven difficult (e.g., Snow, 2002).

Approach 3 is highly analytic and involves providing interactive assistance in the form of comprehension strategies to students engaged in reading content-area passages in computer-based instruction environments. When effective, the third approach offers two complementary forms of outcomes: (a) one that provides greater understanding of the reading comprehension process itself and (b) one that could provide a means for the delivery of effective content-area reading comprehension instruction in school settings. As noted previously, the issue is not which of these settings is best but rather how can research be designed to support their integration with one another in a form that furthers an understanding of the reading comprehension process.

LINKING KNOWLEDGE, EXPERTISE, AND READING COMPREHENSION

In this section, we consider perspectives from applied cognitive science that integrate the role of prior knowledge in meaningful learning (i.e., content-area comprehension) with emphasis on the link between knowledge-based instructional models and the development of general reading comprehension proficiency.

Knowledge-Based Instruction Models as a Foundation for Meaningful Learning

The distinguishing characteristic of knowledge-based instruction models is that all aspects of instruction (e.g., teaching strategies, student learning activities, assessment) are related explicitly to an overall curricular framework that represents the logical structure of the concepts in the subject-matter discipline to be taught. In considering this design characteristic as a key focus for meaningful learning, knowledge-based instruction is best illustrated by the original architecture of computer-based intelligent tutoring systems developed in the early 1980s (e.g., Kearsley, 1987; Luger, 2002). In intelligent tutoring systems, the explicit representation of the knowledge to be learned serves as an organizational framework for all elements of instruction, including the determination of learning sequences, the selection of teaching methods, the specific activities required of learners, and the evaluative assessment of student learning success. In considering the implications of knowledge-based instruction for education, it is important that one of the strongest areas of cognitive science methodology focuses on explicitly representing and accessing knowledge (e.g., Kolodner, 1993, 1997; Luger, 2002; Sowa, 2000). Therefore, the general methodological perspectives that guide knowledge-based educational applications and research should be considered as well established.

A Knowledge-Based Perspective for Considering Comprehension and Reading Comprehension

Although the role of knowledge in meaningful learning (i.e., comprehension) has received some previous notice in education (e.g., Carnine, 1991; Glaser, 1984; Hirsch, 2001; Kintsch, 1998), such attention was minimal until the recent National Research Council publication *How People Learn* (Bransford et al., 2000). In their book, Bransford et al. (2000) offered an informal conceptual overview of the role of knowledge in meaningful learning. In approaching meaningful learning in a fashion equivalent to comprehension, Bransford et al. emphasized consensus research comparing experts and novices in two areas of investigation. The first area summarized research showing that experts displayed greater in-depth conceptual frameworks for organizing their knowledge that, in turn, facilitated their subsequent access and application of knowledge to better understand (i.e., to comprehend) the dynamics of the settings with which they interacted. In contrast, novices commonly attended to irrelevant surface features, using weak organization schemes that did not enhance their comprehension of the dynamics they faced. The second area emphasized the important role of conceptual frameworks as the form of prior knowledge that facilitates new meaningful learning (i.e., comprehension in learning tasks).

An important implication from the Bransford et al. (2000) book supported by a wide variety of sources (e.g., Carnine, 1991; Glaser, 1984, Kintsch, 1998; Vitale & Romance, 2000) is that curriculum mastery can be considered as a form of expertise and, within the limitations (or scope) of what students learn within a content discipline, that mastery of conceptual academic content by students should prepare them to function in a manner consistent with that of experts. In this regard, the in-depth understanding of core concepts and concept relationships can be emphasized as a critical element of general comprehension and, by inference, of reading comprehension as well.

Figure 4.1 illustrates a knowledge-based perspective of reading comprehension that is consistent with the broad idea of meaningful learning presented by Bransford et al. (2000). Figure 4.1 suggests that the nature of comprehension in general learning and in reading settings are equivalent, with the exception that the specific learning experiences associated with reading comprehension are text dependent. With this equivalence in mind, three scenarios for reading comprehension are outlined in the figure. In Scenario 1, the reader is learning an elaboration of prior knowledge, so the new knowledge is incorporated into what is known. In Scenario 1, which represents content-area reading expertise based on the accessibility of domain-specific knowledge, no explicit comprehension strategies are required; however, comprehension as the assimilation of knowledge does imply prior knowledge in the form of a core conceptual framework.

In Scenario 2, the existing framework of prior knowledge is not adequate to assimilate new knowledge, so the reader must identify the new content to be understood and then organize it by meaningfully integrating prior and new knowledge. Thus, Scenario 2 does require metacognitive strategies that, in this chapter, are represented as a coordinated three-part knowledge-focused reading comprehension strategy (i.e., a text elaboration substrategy, a propositional concept mapping substrategy, a summarization/writing substrategy).

Finally, in Scenario 3, the content of the source material in conjunction with prior knowledge is not sufficient for meaningful understanding. Therefore, in Scenario 3 the reader must apply heuristic strategies to obtain the additional knowledge needed and then proceed according to Scenario 2. In Scenario 3, having prior experience in addressing such informational deficiencies (and having access to supplementary sources) is a logical requirement for effective independent learning. As a conceptual model, Scenario 3 is transformed into Scenario 2 by obtaining additional information, after which Scenario 2 is transformed into Scenario 1 through the application of reading comprehension strategies that create a new organizational framework for integrating prior and new knowledge. In turn, Scenario 1 provides a framework for understanding future new knowledge in a form that allows it to function as prior knowledge for new learning.

Together, these three scenarios reflect a form of reading comprehension proficiency that can be considered characteristic of expert readers. As Figure 4.1 suggests, training students to use reading comprehension strategies in a setting

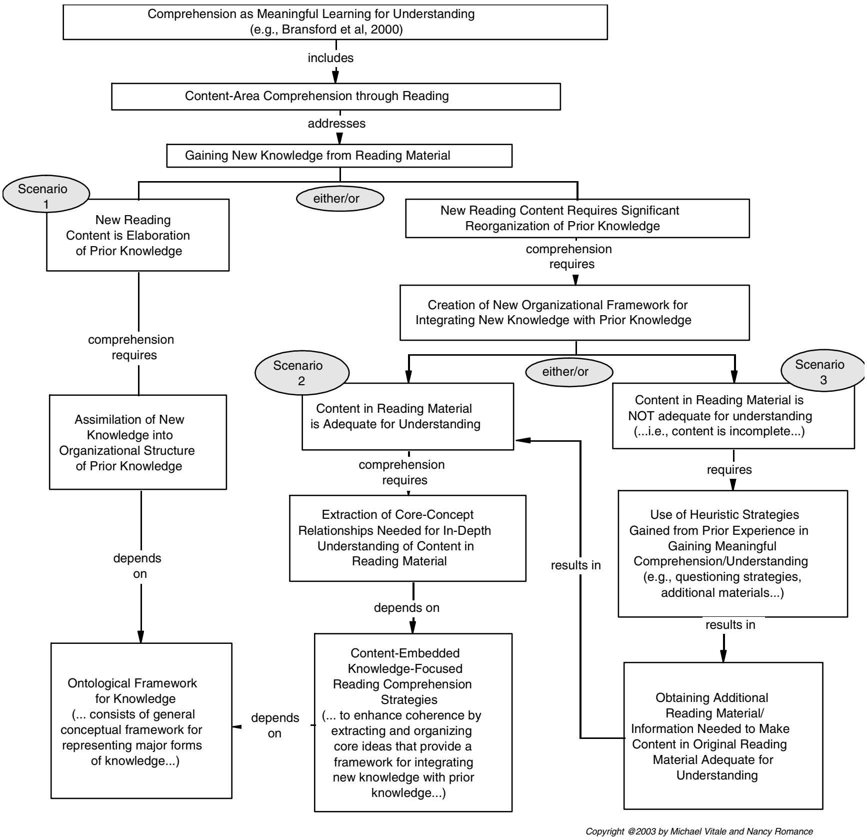


Figure 4.1. A knowledge-based perspective considering reading comprehension as a subset of meaningful understanding. Scenarios 1, 2, and 3 identify different contexts for reading comprehension.

that does not embed them in content-reading applications that require meaningful learning raises the question of ecological validity and, as a result, their subsequent transferability to authentic content-area reading comprehension settings (see Niedelman, 1992). As noted earlier, summaries of research (e.g., Snow, 2002; Trabasso & Bouchard, 2002) have reported that demonstrations of the transfer of reading comprehension strategies from the specific instructional conditions in which they are learned to other applied settings have met with only limited success.

Reading Comprehension as a Skill-Based Proficiency

If comprehension gained through reading can be considered a form of expertise, then it is relatable to general work in cognitive science by Anderson and others (e.g., Anderson, 1982, 1987, 1992, 1993, 1996; Anderson & Fincham, 1994). This work distinguishes the strong problem-solving processes of experts that are highly knowledge based and automatic from the weak strategies of novices having minimal knowledge that may range from heuristics to trial-and-error search. Within the context of reading comprehension, use of the prior content (or world) knowledge that students bring to reading tasks can be considered a form of knowledge-based problem solving, whereas reading comprehension strategies can be considered to serve as weak problem-solving strategies (i.e., as metacognitive or heuristic tools) that, when well developed, become automatic. In reading comprehension tasks, such strong and weak processes presumably operate in a complementary fashion at a level of automaticity for expert readers in both general and reading comprehension learning tasks. Extrapolating from the just-cited work by Anderson and others, the consideration of reading comprehension strategies to be cognitive skills implies that they require extensive amounts of varied practice to reach the degree of automaticity that is characteristic of expert performance.

In related work, both Niedelman (1992) and Anderson and colleagues (e.g., Anderson, 1996) have offered interpretations of research issues relating to transfer of learning that are consistent with a knowledge-based approach to learning and understanding and that are directly applicable to reading comprehension. Such work on transfer of learning is of major importance for research pursuing an understanding of the potentially differential effects of having students learn to apply reading comprehension strategies when they are embedded within (i.e., operate on) a content-domain involving cumulative, meaningful learning (e.g., science, history, geography) as opposed to when such strategies are applied in noncumulative learning settings (e.g., narrative stories in typical basal reading programs) that are different from the content-oriented contexts to which use of such strategies are to be transferred. In this regard, a comprehensive interdisciplinary review of reading comprehension research by McNamara, de Vega, and O'Reilly (in press) concluded that skilled comprehenders are more able to actively and efficiently use knowledge (and strategies) to help them comprehend text and, furthermore, that individual differences in reading comprehension depend on the dynamics associated with knowledge activation.

These research perspectives, in conjunction with Figure 4.1, also suggest that student variability in reading comprehension proficiency could be considered to reflect one or more of the following three dynamic factors operating in a complementary fashion: (a) the development of domain-specific prior knowledge, (b) the development of reading comprehension strategies as metacognitive tools, and (c) the development of heuristic strategies for obtaining additional sources of information. These dynamics presumably operate automatically for expert readers (i.e., self-reliant readers) across different contexts. It is important to note that each of these three key dynamics is amenable to instruction.

Application of a Knowledge-Based Instruction Model for Building General Reading Comprehension

Science IDEAS is a research-based, cognitive science oriented, instructional intervention that was initially validated within a Grade 4 upper elementary setting (Romance & Vitale, 1992). As a knowledge-based instruction model, Science IDEAS combines science, reading, and writing through daily 2-hr time blocks that replace regular reading/language arts instruction. Multiday science lessons in Science IDEAS engage students in a variety of instructional activities (e.g., reading text/trade/Internet science materials, hands-on science experiments, writing about science, journaling, propositional concept mapping as a knowledge representation tool), all of which focus on the cumulative enhancement of science concept understanding. As an instructional intervention implemented within a broad inquiry-oriented framework (i.e., all aspects of teaching and learning emphasize learning more about what is being learned), teachers use science concepts and relationships (which students master to develop in-depth science understanding) as curricular guidelines for identifying, organizing, and sequencing all instructional activities.

The effectiveness of the Science IDEAS intervention in engendering reading comprehension proficiency is well established. The initial research study implemented in Grade 4 classrooms (Romance & Vitale, 1992) found that, compared with demographically similar controls, the Science IDEAS model resulted in significantly higher levels of student achievement on nationally normed tests not only in science (adjusted mean difference in Metropolitan Achievement Test Science = .95 GE) but also in reading comprehension (adjusted mean difference in Iowa Tests of Basic Skills [ITBS] Reading Comprehension = .32 GE). Science IDEAS students also displayed significantly more positive attitudes toward science learning, more positive self-confidence in learning science, and more positive attitudes toward reading.

Using the initial findings as a foundation, the Science IDEAS intervention subsequently was extended to a greater number of classrooms in Grades 3 through 5 over a 3-year period involving more than 50 teachers and 1,200 students, including ethnically diverse student populations and a variety of academic levels, ranging from above average to severely at risk. As we reported (Romance & Vitale, 2001), the findings revealed a similar and consistent pattern of positive learning effects on norm-referenced tests in both science and reading comprehension (along with similar affective outcomes). In addition, the Year 4 study addressed an important equity issue by showing that there were no differences in rate of achievement growth and in affective outcomes due to student demographics (e.g., at-risk status, gender, race).

Of importance to this chapter is the finding that a knowledge-based, conceptually oriented, science intervention (Science IDEAS) in which teachers did not instruct students on reading comprehension strategies that research has recognized as important (e.g., Gersten et al., 2001; Trabasso & Bouchard, 2002) obtained consistently higher achievement in reading comprehension than alternative (basal) reading curricula whose purpose was to teach reading comprehension. Furthermore, because the specific (mostly nonscience) content of the reading passages of the nationally normed reading comprehension tests used in the studies (ITBS, Stanford Achievement Test) were different than the specific science content students learned and read about in their classrooms, the positive effect of Science IDEAS clearly represented a general transfer of learning outcome in reading comprehension (see Niedelman, 1992).

Although a variety of interpretations are consistent with the transfer effects from Science IDEAS to general reading comprehension, one is that knowledge-based learning involves combinations of points made by Bransford et al. (2000) that emphasize the importance of the development of prior knowledge in meaningful learning, the work of Kolodner (e.g., 1993, 1997) on case-based knowledge representation and reasoning, the factors relating to the development of knowledge-based expertise summarized by Bransford et al. and Anderson (1996), and the general ontological functions of knowledge representation offered by Sowa (2000). Although speculative, this interpretation is suggestive of implications for future research for investigating the dynamics of how students gaining cumulative in-depth understanding of science resulted in the development of general reading comprehension proficiency. One possible starting point for such research is the idea that cumulative, meaningful, content-area learning results in the developmental refinement of a general framework (see Vitale & Romance, 2000, 2006a, 2006b) for using core concepts and concept relationships that engender expertise-based new learning. Within a knowledge-based framework, this interpretation is consistent with Sowa's (2000) analysis of the ontological functions of knowledge representation and the complementary work of both Anderson (1996) and Sidman (1994, 2000) emphasizing the importance of extensive and varied practice in the development of concepts and concept relationships that were formal instructional characteristics of the Science IDEAS model (see also Mintzes, Wandersee, & Novak, 1998).

Related Research in Reading and Educational Psychology Focusing on Reading Comprehension

Consistent with the preceding knowledge-based perspective is a substantial body of literature in the area of reading comprehension. In an extensive summary of text comprehension strategy instruction, Trabasso and Bouchard (2002) examined 205 empirically based studies of 12 distinct cognitive strategies for improving reading comprehension (e.g., comprehension monitoring, graphic organizers, prior knowledge, question generation, story structure, summarization, vocabulary instruction) that were conducted from 1980 through the review date. In their conclusions, they emphasized the importance of episodic content knowledge as a basis for reader-constructed deeper understanding, the related use of graphically oriented story mapping (see also Williams, 2002) as a means for guiding student explication of narrative understanding and the related role of student summarization involving identification and organization of core concepts and themes in material that is read.

Among the most important findings reported by Trabasso and Bouchard (2002) was that the use of multiple-strategy instruction taught through dialogue-rich teacher modeling/guidance was a powerful approach for improving student reading comprehension proficiency. In identifying directions for future research, they emphasized the importance of conducting reading comprehension strategy research in content-area instruction and in focusing such research on the issue of enhancing the transferability of reading comprehension strategies. In a complementary review, Gersten et al. (2001) reported similar conclusions (see also Farstrup & Samuels, 2002).

In a related review focusing on children's searching and using informational text, Dreher (2002) stressed the importance of substantially expanding the instructional experiences of upper elementary students with informational (content-oriented) text. Ogle and Blachowicz (2002) presented similar concerns relating to the need to emphasize informational text at the elementary levels. In a review of research designed to improve the comprehension of expository text, Pearson and Fielding (1995) found that organizational enhancements, such as summarizing text structure (e.g., hierarchical elaboration summaries, visual organizers), were powerful in facilitating overall comprehension and learning. Finally, within a context of discourse analysis, Weaver and Kintsch (1995) noted the importance of the structure of domain-specific prior knowledge in affecting how text is understood and remembered in general and how the interactive nature of domain-specific knowledge affects the effectiveness of reading comprehension strategies in particular (see also Perkins & Grotzer, 1997).

Although referenced in reviews cited earlier, the extensive work by Guthrie and his colleagues (e.g., Guthrie et al., 1998; chap. 10, this volume; Guthrie & Ozgungor, 2002; Guthrie, Wigfield, Barbosa et al., 2004; Guthrie, Wigfield, & Perencevich, 2004) is important to recognize. This work has shown repeatedly that engaging upper elementary students with content-oriented reading materials (e.g., science, social studies) has a significant effect on both reading proficiency and student motivation to engage in reading. In this regard, Armbruster and Osborn (2001) summarized research findings demonstrating positive student achievement in reading comprehension resulting from integrating science content with reading/language arts. Finally, other sources (Beane, 1995; Ellis, 2001; Hirsch, 1996; Schug & Cross, 1998; Yore, 2000) have discussed issues and findings that support interventions in which core curriculum content is used as a framework for addressing reading comprehension.

DESIGN OF KNOWLEDGE-FOCUSED READING COMPREHENSION STRATEGIES

In this section, we consider the development of reading comprehension strategies from two complementary views. The first suggests criteria based on the research literature as guides for the design of reading comprehensive strategies that are effective in school settings. The second presents a multipart knowledge-focused reading comprehension strategy based on the criteria, along with the results of a demonstrational (i.e., proof-of-concept) study (Romance & Vitale, 2005) designed to explore the potential effectiveness of a multipart strategy within content-oriented (science) and narrative-oriented (basal reading) instruction in Grade 5 school learning settings.

Toward Research-Based Criteria for Reading Comprehension Strategy Design and Analysis

An important question for researchers and practitioners is how to optimize the effectiveness and usability of reading comprehension strategies with the forms of cumulative meaningful learning that occur in applied school settings (vs. more short-term controlled research settings). As noted previously, this concern was identified as a major problem in the RAND report (Snow, 2002). In fact, questions of effectiveness and usability are the two major considerations that logically underlie the design of any form of instructional intervention (see Dick, Cary, & Cary, 2004).

With the preceding in mind, three interdependent criteria for reading comprehension strategy design and analysis can be identified. The first criterion is *power*. Following experimental approaches for establishing the generalizability of research findings outlined by Sidman (1960), the idea of power is that effective reading comprehension strategies must address major sources of variance in learning if they are to have a replicable (and dynamic) impact across applied settings within which many other types of uncontrolled learning dynamics are operative.

The second criterion is *functionality*. The idea of functionally is that any reading comprehension strategy should be proceduralized in a fashion that explicates its enhancement of the ongoing process of reading comprehension in a causal fashion. That is, use of the strategy should enhance comprehension explicitly rather than require comprehension as a prerequisite for use (i.e., a strategy that requires comprehension has a testing rather than a tool function).

The third criterion is *engineering efficiency*. The idea of engineering efficiency is that the reading comprehension strategy should be usable easily by teachers and students. Such engineering efficiency occurs most readily when a strategy results from an explication of naturally ongoing processes (e.g., teacher reading expertise in this study) that result from the analytic identification of the key elements of such processes. In turn, the key elements that define such processes can be implemented through standard forms of training that result in their use as a strategy that can ultimately be applied by learners (i.e., through a sequence of modeling, guiding, and independent practice of the strategy). In particular, the idea of using an expertise-oriented approach for designing reading comprehension strategies to be used by teachers meets this requirement (i.e., teachers are able to reflect on their own behavior as a guide for using the reading comprehension strategy and then to use their reflections as a guide for modeling and supporting student use of the strategy).

Considered together, the three criteria provide a strong set of constraints for either designing or analytically evaluating reading comprehension strategies. With regard to the design of the multipart reading comprehension strategy described in this chapter, the criterion of power applies in that the substrategies focus on amplifying the access of prior knowledge for comprehension of what is being read and then organizing and integrating new knowledge with existing knowledge in a form that is accessible as prior knowledge for future comprehension. The criterion of functionality applies in that each of the substrategies is described as a procedure whose application enhances the process of comprehension, both in real time reading and in follow-up reviews of what has been read. Also, the criterion of engineering efficiency applies in that the substrategies are explications of the forms of expertise all teachers display naturally when they read with comprehension and on whose use students can be guided as they read or reflect on what they have read in applied learning settings.

In a related fashion, the three criteria also could be used to analyze the potential effectiveness of reading comprehension strategies typically used by practitioners. In general, the majority of such strategies (see Billmeyer & Barton, 1998, p. 69, for an extensive list) either require comprehension as a prerequisite for use (i.e., confuse engendering comprehension when reading with testing for comprehension after reading) or address aspects of comprehension that are of minimal importance to content-area understanding. Although the majority of the 40 different strategies reported by Billmeyer and Barton (1998) are easy to use, such ease of use does not matter if the strategies are inadequate for substantially enhancing reading comprehension as a form of understanding when used in classroom settings.

A Multipart Reading Comprehension Strategy That Meets the Design Criteria

Using the preceding criteria as a foundation, in this section we outline a multipart knowledge-focused reading comprehension strategy consisting of three complementary substrategies that together explicitly address (and support) research-recognized aspects of the process of reading comprehension. In describing these substrategies (text elaboration substrategy, propositional concept mapping substrategy, summarization/writing substrategy), it is important to note that each represents and augments aspects of the process of reading with comprehension as a form of expertise and addresses the key roles of knowledge in reading comprehension (i.e., knowledge access and organization). As forms of expertise, the substrategies are engineered for use by teachers in regular classroom settings and have a strong research base (see Dreher, 2002; Snow, 2002; Trabasso & Bouchard, 2002).

Text Elaboration Substrategy. The teacher planning and implementation procedures for the text elaboration substrategy are summarized in Table 4.1. This is the pivotal substrategy among the three, because it operates in real time and forms a basis for the complementary use of the other substrategies. As shown in Table 4.1, the text elaboration substrategy is a reflective explication of the prior knowledge individual teachers think about as they read a passage with comprehension (i.e., related knowledge that helps make the passage meaningful). In a parallel fashion, the implementation of the substrategy consists of teachers guiding students to actively relate what they are reading both to their prior knowledge in general and to what they previously have read with understanding in the passage itself. Through rereadings of the passage,

TABLE 4.1

Overview of Text Elaboration Sub-Strategy

Teacher Planning Process :

- Read passage for understanding
- Re-read and generate "knowledge notes"
  - Read sentence or set of sentences
  - Think about what knowledge made the text passage understandable (also flag key ideas)
- Link "knowledge notes" (via post-its) to passage locations
- Transform "knowledge notes" into "knowledge link" questions answered by the "knowledge notes" (for use with students)
- Result of teacher planning process
  - Teacher has meaningful set of questions for guiding student use of prior knowledge for comprehension
  - Questions are based on individual "teacher expertise" in reading passage with understanding

Teacher Implementation Process

- Initial reading of passage
  - Select student to read passage/section aloud
  - During reading, model (ask/answer), guide (ask) "knowledge-link" questions, or accept student-initiated "knowledge links," as appropriate
  - Guide:
    - student passage/section summary
    - student cumulative passage summary
  - Continue process with new students until passage read
- Multiple re-readings of same passage
  - Select student to read passage/section
  - During reading, guide or solicit/accept student-initiated "knowledge-link" questions/answers as evolving process
  - Solicit:
    - Student passage/section summary
    - Student cumulative passage summary
  - Continue process with new students until passage re-read
  - repeat re-reading until students are able to initiate "knowledge links"

Note 1: Teachers model, guide, accept student-initiated "knowledge link" questions/answers as an evolving process (objective is to obtain student-initiated knowledge-links across repeated re-readings of the same passage and then generalize the process across new passages)

Note 2: Teacher knowledge-link questions emphasize linking what is being read to what has been read previously and to prior knowledge

Note 3: Goal of sub-strategy is for students to learn to use prior knowledge for reading comprehension

TABLE 4.2
Overview of Propositional Concept Mapping Sub-Strategy

Teacher Planning Process:
- Read passage for understanding
- Identify key ideas/examples and write on post-it notes
- Organize ideas/examples in hierarchical structure (via post-it notes), arranging big ideas on top, sub-ideas below, and examples on bottom
- Generate links for connecting concepts so each concept-link-concept unit is in form of a simple sentence (i.e., noun-verb-noun)
- Result of Teacher Planning Process:
  - Teacher has coherent organizational structure representing the core knowledge in text passage
  - Structure can be used in a variety of ways (e.g., planning instruction/assessment), but emphasis here is on planning for students learning to concept map

Teacher Implementation Process:
- Students read passage and identify key ideas/examples, write ideas/examples on post-it notes
- Students identify core ideas, subordinate ideas, examples, arranging post-it notes in real-time to form hierarchical structure
- Students identify links that form concept-link-concept units into simple sentences (i.e., noun-verb-noun)
- Students read the map as if it were prose (editing/re-reading as necessary)

Note 1: Teachers model, guide, accept student-initiated participation as appropriate (i.e., as evolutionary process)

Note 2: Goal of sub-strategy is for students to learn to organize/represent knowledge into a coherent structure to enhance comprehension and accessibility

knowledge that appears later in the passage can be related as prior knowledge to earlier parts of the passage previously read, a sequence that is particularly useful with technical content. Overall, this substrategy is designed to teach students to actively relate what they are reading to prior knowledge gained through previous learning, including the preceding content in the passage being read, as a means to enhance comprehension.

Propositional Concept Mapping Substrategy. The teacher planning and implementation procedures for the propositional concept mapping substrategy are summarized in Table 4.2. As shown in the table, the propositional concept mapping substrategy is best understood as a follow-up to the application of the text elaboration substrategy to a specific passage (or series of passages) in which students identify, arrange in hierarchical form, and link together the core ideas, subordinate ideas, and illustrative examples in a visual display (see Figure 4.2 for an example of a propositional concept map). Again, in a

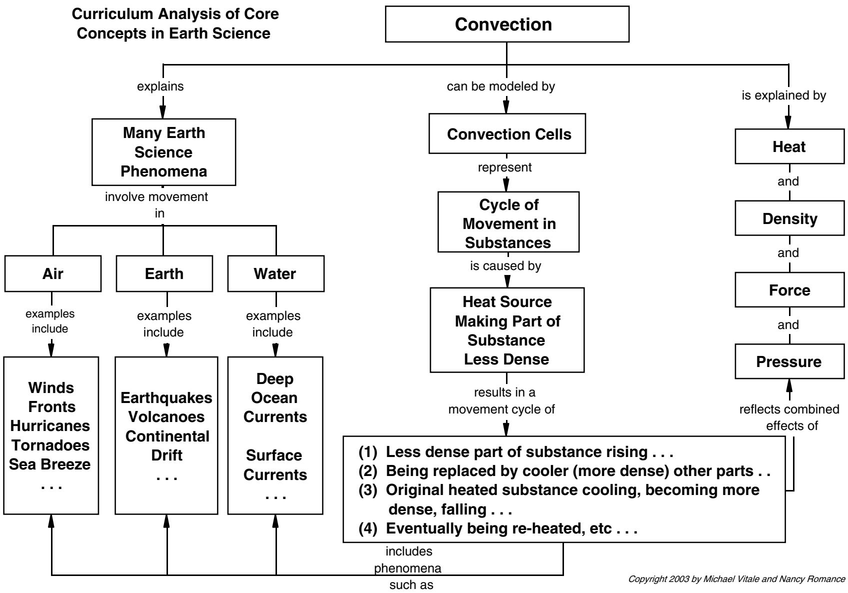


Figure 4.2. Example of a propositional concept map in earth science showing the role of convection.

fashion paralleling that of the text elaboration substrategy, the implementation of the concept mapping substrategy as a process consists of teachers guiding students to build the propositional concept map as a meaningfully organized representation of the knowledge in passages they have read. Overall, the concept mapping substrategy is designed to teach students to actively organize the knowledge about which they are reading by identifying key concepts and concept relationships to enhance comprehension.

Summarization/Writing Substrategy. The teacher planning and implementation procedures for the summarization/writing substrategy are summarized in Table 4.3. As shown in the table, the summarization/writing substrategy is primarily a side effect of the propositional concept mapping substrategy that involves transforming propositional concept maps into standard text prose, a task for which the structure of propositional concept maps provides strong prosthetic support. Overall, this substrategy is designed to teach students to meaningfully access and organize knowledge they want to communicate as a prerequisite for maximizing their writing coherence.

TABLE 4.3
Overview of the Summarization/Writing Sub-Strategy

Teacher Planning Process (None: Side-effect of concept mapping):
Teacher Implementation Process:
– Students use organizational structure of propositional concept map as a guide for written summary of passage (parallels process of reading the map in concept mapping)
– Edit (or elaborate) written summary as appropriate

Note. Goal of sub-strategy is for students to learn to develop and then access an organized knowledge structure as a basis for coherent writing.

Operational Aspects of the Knowledge-Focused Reading Comprehension Strategy. From an operational standpoint, the three substrategies comprising the knowledge-focused reading comprehension strategy are designed to be used in a complementary fashion on a continuing basis across a wide variety of reading comprehension tasks. Together, the combined focus of the three substrategies provides students with the means to approach such tasks as reading occurs in a fashion that enhances in-depth understanding of the knowledge to be gained by accessing prior knowledge, representing/organizing new knowledge learned with prior knowledge, and generating oral and written expressions that summarize understanding of knowledge in a coherent fashion. Applied across cumulative, meaningful learning environments, the potential result is the integration of new and existing knowledge as a form of expertise. Additionally, the engineering design also allows the substrategies to be learned and applied by students as a form of expertise that consists of an elaborative transfer of naturally occurring processes that engender comprehension.

Exploring the Effectiveness of the Reading Comprehension Strategy in Content Versus Narrative Instruction

Research Context. In this section, we summarize a proof-of-concept study (Romance & Vitale, 2005) that addressed two complementary elements. The first element consisted of the findings reported by Snow (2002) regarding the lack of success in the field of reading regarding the improvement of content-area reading comprehension and the corresponding lack of evidence that the effects of reading comprehension strategies in experimental settings are transferable to applied settings. The second element consisted of the findings (Romance & Vitale, 2001) that showed that in-depth science instruction over a school year within which reading and language arts were integrated (Science IDEAS) was more effective than traditional reading instruction in engendering reading comprehension proficiency as measured by nationally normed tests.

Overall Study Design. The study had three specific goals. The first was to determine whether the multipart knowledge-focused reading comprehension strategy was effective in enhancing student proficiency in reading comprehension and achievement in science. The second was to determine whether the strategy was differentially effective in two instructional settings: (a) one in which reading was integrated within content-oriented science instruction (Science IDEAS) and (b) one in which reading instruction was implemented within a traditional basal reading program. The third goal was to replicate our previous findings (Romance & Vitale, 1992, 2001) that found that the Science IDEAS model resulted in greater student achievement in reading comprehension than traditional basal reading/language arts instruction.

In implementing the multipart knowledge-focused reading comprehension strategy as a composite practitioner-oriented intervention model, the study adopted a design structure for programmatic research in applied school settings presented by Slavin (1990, 2002). In distinguishing between variable-oriented analytic studies preferred by researchers and model-oriented applications useable by practitioners, Slavin noted that both types of studies could be pursued efficiently by first demonstrating the effects of a practically oriented composite model in school settings (e.g., the multipart reading comprehension strategy in this study) and then conducting analytic, variable-oriented, follow-up research to optimize the different elements of the composite model.

Classrooms that used the reading comprehension strategy intervention also engaged students an oral semantic fluency activity (Vitale & Medland, 2004) as a methodological enhancement to facilitate student verbal classroom responsiveness to the reading comprehension strategies. Although there is an extensive research literature on using semantic (or ontological) frameworks as an analytic comprehension tool (e.g., O'Donnell, Dansereau, & Hall, 2002), this study did not link the oral semantic fluency activities to either the specific academic content to be learned or to the reading comprehension strategy activities themselves. Because its role in the study was methodological, the semantic fluency activity is not considered further here.

Student reading comprehension proficiency and science achievement were assessed on a pre- and posttest basis. Pretests were part of the state-administered Florida Comprehensive Assessment Test Battery (FCAT) in reading (Stanford Achievement Test Reading Comprehension) and science (FCAT Science Assessment Test) that occurred before the start of the study in mid-March 2004 and served as covariates for statistical control in data analysis. Posttests that served as criterion measures of student achievement were nationally normed reading (ITBS Reading Comprehension, Level 11) and science (ITBS Science, Level 11) tests administered at the end of the study, in mid-May 2004.

Implementation. The study was implemented on a schoolwide basis in Grade 5 in a total of 6 demographically representative elementary schools located in a large, diverse school system in southeastern Florida. Four of the participating schools used Science IDEAS, and 2 used a district-adopted traditional basal series. The 4 Science IDEAS schools were selected randomly from a total of 11 schools implementing the model, whereas the 2 traditional schools were selected from a pool of demographically comparable schools that volunteered for participation in the study. The reading comprehension strategy intervention was assigned on a random basis to two of the four Science IDEAS schools and one of the two traditional basal schools. Because schools in the study previously had selected either Science IDEAS or the district-adopted basal reading series on a voluntary basis (i.e., volunteering for participation in Science IDEAS for the school year was an option for all district schools), the selection of the demographically similar non-Science IDEAS schools was considered an adequate methodological control for the study.

The experimental interventions used in the 8-week study are outlined in Figure 4.3. Both Science IDEAS (Romance & Vitale, 2001) and the district-adopted basal reading/language arts program, the Scott-Foresman Reading for Florida (Scott-Foresman, 2002), were implemented through daily 2-hr time blocks. In accordance with district curricular time allocations, participating Science IDEAS schools allocated a limited amount of additional time (approximately 30 min/day several days per week) to address literary content, whereas Scott-Foresman schools allocated an equal amount of additional time on a weekly basis to teach science.

In Science IDEAS schools, the knowledge-focused reading comprehension strategy was implemented as described above. However, as shown in Figure 4.3, the propositional concept mapping substrategy was adapted to function as a story mapping substrategy that was more appropriate for the schools using the traditional basal reading program, which emphasized narrative rather than content-oriented reading. Because the story mapping activity used an explicit graphic framework, it was closer in operation to structural analytic frameworks (e.g., O'Donnell et al., 2002) used to enhance reading comprehension than to the relatively free-form propositional concept mapping used in Science IDEAS instruction. However, in this study both variants represented visual/graphic enhancements of the elements identified by students applying the text elaboration substrategy to their respective reading materials.

In each of the two instructional settings, the substrategy procedures were modeled and guided for students by teachers until students were able to engage in each substrategy procedure. Thus, all three substrategies were applicable to informational and narrative text material and reflected key characteristics of

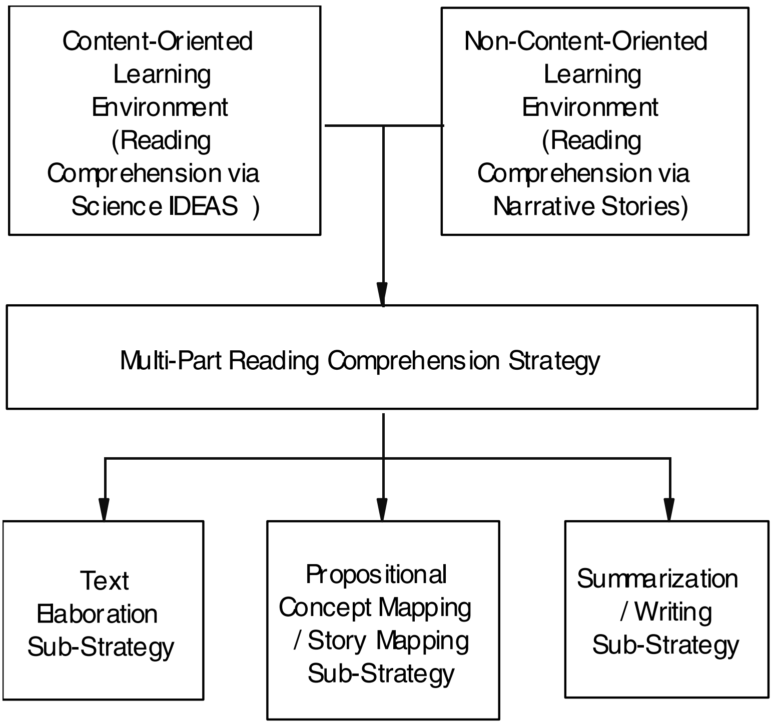


Figure 4.3. Elements of the two-factor experimental intervention (type of instructional environment and use of the knowledge-focused reading comprehension strategy).

expert readers in those settings. Teachers used the text elaboration substrategy once each week and the propositional concept mapping/story mapping substrategy and summarization/writing substrategy once every 2 weeks. Both sets of teachers received initial and continuing support for implementing the reading comprehension strategies in the study.

Results. As a group, the students who received the experimental (knowledge-focused reading comprehension strategy) performed similarly to the control students on the FCAT/Stanford Achievement Test Reading and FCAT Science premeasures (unweighted mean pretest differences between experimental and control schools = 4.0 scale units in reading and 6.7 in science), although the school selection process considered only schoolwide demographic factors (i.e., free/reduced lunch, student ethnicity) because FCAT results were not available before the start of the study. With respect to Science IDEAS versus traditional basal reading schools, the unweighted mean differences were 30.6 and 34.8 in favor of Science IDEAS schools for FCAT/Stanford Achievement Test Reading and FCAT Science, respectively. The difference in reading was due to the withdrawal of a higher achieving school just before the initiation of the study (too late to be replaced). However, the pretest differences in reading were controlled statistically through the linear models analysis methodology. The pretest difference in science achievement (which was less than 1 standard deviation) was considered statistically controllable as well.

The results of the multivariate covariance analysis implemented through a general linear models approach found that the main effect of content-oriented (Science IDEAS) versus non-content-oriented (traditional basal) instructional environment was significant, with the Science IDEAS classrooms performing significantly higher in both reading, $F(1, 557) = 7.29$, $p < .01$ (adjusted mean difference = .38 GE for ITBS Reading) and science, $F(1, 557) = 4.84$, $p < .05$ (adjusted mean difference = .42 GE for ITBS Science). Although the main effect of use of the reading comprehension strategy treatment was not significant, a significant interaction between use of the reading comprehension strategy and instructional environment was found for both reading, $F(1, 557) = 85.70$, $p < .01$, and science, $F(1, 557) = 59.75$, $p < .01$. Follow-up simple effects analysis found that reading comprehension strategy use significantly improved achievement for both ITBS Reading, $F(1, 557) = 6.63$, $p < .01$ (adjusted mean difference = .53 GE) and ITBS Science, $F(1, 557) = 4.49$, $p < .05$ (adjusted mean difference = .17 GE) for the science-oriented instructional environment (Science IDEAS) students. However, use of the reading comprehension strategy was not significant for either ITBS Reading or Science in the traditional reading/language arts environments that emphasized narrative reading.

Discussion. The results of the study are supportive of the major arguments presented in this chapter. First, the content-oriented Science IDEAS model resulted in greater reading comprehension achievement overall than did traditional basal reading instruction. This finding was consistent with those we reported (Romance & Vitale, 2001) and reaffirmed the principle that engaging students in in-depth, cumulative, meaningful learning is a more effective approach for developing reading comprehension proficiency than repeatedly engaging students in unrelated stories that are designed to eliminate cumulative content-area knowledge (see Hirsch, 1996; Walsh, 2003). Second, the fact that the use of the reading comprehension strategy enhanced reading comprehension only for Science IDEAS students is suggestive that being engaged in content-oriented learning environments that emphasize cumulative meaningful learning may be an important condition for positive transfer of the effects of metacognitive instructional enhancement procedures such as reading comprehension strategies. This interpretation is consistent with Snow's (2002) report that the well-established research base in reading (e.g., Block & Pressley, 2002) has found only minimal evidence of the positive influence of reading comprehension strategies to content-area reading in applied settings. Third, the findings are consistent with the broader perspective that reading comprehension can be approached as a special case of comprehension in general. In this regard, replication and extension of this study should be undertaken to explore the interactive treatment effects found in the present 8-week study over longer instructional time intervals in applied school settings.

A primary question raised earlier in this chapter was why in the original studies we reported (Romance & Vitale, 2001) resulted in Science IDEAS students displaying greater reading comprehension achievement than students in traditional reading/language arts instruction, given that Science IDEAS students received no formal instruction in reading or guidance in use of reading comprehension strategies. Although teachers who implemented the original Science IDEAS model did not use explicit reading comprehension strategies, it is clear retrospectively that Science IDEAS teachers did incorporate processes similar to those used in the multipart knowledge-focused reading comprehension strategy in this study by relating different types of activities, including different readings, that focused on the same science concepts. However, in doing so they focused generally on the meaningful (overall) comprehension of science content than on reading comprehension per se in producing reading comprehension proficiency transfer results.

For example, in studying a topic such as evaporation, students would actively relate different kinds of activities (e.g., hands-on experiments, reading, concept mapping, journaling/writing, projects) to the common set of concepts and concept relationships being studied. So, as students engaged in reading, they were actively involved in relating what they were reading to prior knowledge they would have gained through earlier reading and other activities. Insofar as this study is concerned, the processes for in-depth learning in Science IDEAS and the knowledge-focused reading comprehension strategy can be considered variants of a common instructional architecture (see Dillon & Tan, 1993) for developing content-area comprehension by relating what is being learned to prior knowledge (see also Gagne, Wager, Golas, & Keller, 2004; Vitale & Romance, 2006a, 2006b). In this regard, the more explicit use of the knowledge-focused reading comprehension strategy might be expected to magnify the effects of the Science IDEAS instructional model, as was found in this study.

INTERDISCIPLINARY LINKAGE OF THIS CHAPTER TO OTHER RESEARCH AREAS

In presenting perspectives in a fashion meaningful to researchers and practitioners, the literature we have cited focuses primarily on general sources in applied cognitive science and reading comprehension research. However, all of the perspectives presented in this chapter are interdisciplinary in that they reflect important work in a number of related areas. In this section, we briefly note selected works in some disciplines with the recognition that the scope and implications of the work cited are far broader than can be considered here.

One important area is the construction-integration model developed by Kintsch and his colleagues (e.g., Kintsch, 1988, 1994, 1998, 2002, 2004, 2005; see also chap. 1, this volume), which has been applied extensively to reading comprehension. Kintsch's model explains the process of reading comprehension by distinguishing between the *propositional structure* (i.e., explicit semantic meaning) of a text that is being read and the prior knowledge the reader brings to the process of reading. Within this context, meaningful comprehension results when the propositional structure of the text is joined with the prior knowledge of the learner. If the propositional structure of the text is highly cohesive (i.e., knowledge is explicitly well represented in propositional form), then there is less demand on the reader's prior knowledge. If the text is not cohesive (i.e., contains significant semantic gaps), however, then the reader's prior knowledge is critical for coherent understanding. In either case, comprehension results from the integration of the propositional structure of the text (textbase) with the reader's prior knowledge and is represented semantically in propositional form (a situation model). Within this framework, much of the research conducted by Kintsch and his colleagues has focused on the interplay of meaningful text structure and the prior knowledge of the reader considered as a learner.

Considered here, the focus of Kintsch's work is directly relatable not only to the general concept of meaningful reading comprehension but also to comprehension in general for cases in which learning experiences are other than text based. As a result, the potential scope of Kintsch's model is far broader than reading comprehension because it potentially addresses the complementary roles of the learner's prior knowledge and the general informational structure of learning environments. In this regard, Kintsch's model fits well with strong instructional models for curriculum development (e.g., Engelmann & Carnine, 1991).

A second area considered here is the work of Landauer (2002) and his colleagues (Landauer & Dumais, 1996, 1997; Landauer, Foltz, & Laham, 1998) addressing vocabulary acquisition as an inductive process that is based on the use of prior conceptual knowledge derived from the experience of the learner. As an element in their research, Landauer et al. (1998) developed a computational model, latent semantic analysis, through which the relatedness of words to words, words to prose, and prose to prose can be expressed mathematically by means of a common statistical index across a large number of underlying semantic dimensions. In effect, Landauer and his colleagues have been able to show that the gist of the meaning of a passage is measurable as a sum of its composite words and, in turn, that the meaning of individual words can be measured in terms of their relationship to other words (e.g., think informally of the meaning of a word as a bag of words, i.e., a group of semantically related words consisting of synonyms and antonyms to which it is related by means of a set of fundamental underlying dimensions). As considered here, Landauer et al.'s (1998) model is directly related to the question of how, once gained, prior knowledge can be measured as a form of comprehension and then how such prior knowledge can function as a foundation for future learning. Again, as with Kintsch's, Landauer's work can be broadened to encompass learning experiences beyond those limited to text that result in comprehension (see, e.g., Landauer, McNamara, Dennis, & Kintsch, in press).

Some other important areas of research also complement the perspectives presented in this chapter. Again, although far broader than the scope of this chapter, Graesser and his colleagues (Graesser, Leon, & Otero, 2002; Graesser, McNamara, & Louwerse, 2003; Graesser, Olde, & Klettke, 2002; Graesser & Wiemer-Hastings, 1999) have applied principles from discourse analysis to narrative and expository reading comprehension. Engelmann and Carnine (1991) presented an instructional design model consisting of principles in algorithmic form for structuring learning experiences so that they are optimally coherent to learners and maximize transfer of learning (see Adams & Engelmann, 1996). Finally, Sidman (1994, 2000) and others (e.g., Artzen & Holth, 1997; Dougher & Markham, 1994) have explored the conditions under which learning outcomes that are not taught can arise indirectly (i.e., inductively) from the structural properties of knowledge learned through instruction. Combining the views in this section with those presented earlier in this chapter provides a rich source of research perspectives for advancing the role of knowledge in comprehension and the relationship between knowledge and the conditions under which reading comprehension strategies can enhance student content-area comprehension.

ISSUES RELATING TO RESEARCH ON READING COMPREHENSION STRATEGIES

A general question raised by this chapter is whether research on reading comprehension strategies can be conducted with ecological validity without adopting a contextual approach that frames such research within the forms of content-area learning environments (e.g., science, history) that require cumulative meaningful comprehension. In approaching this question, the complementary approaches to research in reading comprehension presented in an earlier section are important to consider. With regard to Approach 1, the use of cumulative content-area learning environments to engender reading comprehension proficiency, we are engaged in an ongoing research initiative to replicate and extend the implications of the preceding findings and perspectives (Vitale & Romance, 2006a). Included within the scope of this work is the use of a longitudinal design to assess the effects of the knowledge-focused reading comprehension strategy across Grades 3 through 5 on both direct measures of reading comprehension proficiency and on transfer measures at the 3 through 8 levels. In a related fashion, the work of Guthrie and his colleagues on enhancing reading comprehension proficiency through content-area reading at the upper elementary levels also should be noted (e.g., Guthrie, Wigfield, Barbosa, et al., 2004; Guthrie, Wigfield, & Perencevich, 2004).

At the same time, even if the findings of the content-oriented research of Approach 1 emphasizing in-depth cumulative learning are found to be consistent with and extend prior research (e.g., Romance & Vitale, 2001, 2005) by demonstrating a significant positive impact on student reading comprehension, such teacher-delivered interventions in regular school settings are not methodologically appropriate for analyzing the detailed mechanics of how reading comprehension strategies are supportive of content-area reading comprehension. In this regard, Approach 3 (noted earlier), which uses computer-based implementations as interactive instructional media, is ideal for such analytic work (see McNamara, 2006; chap. 16, this volume; Dalton, Palincsar, Defrance, & Hapgood, 2006). However, for purposes of ecological validity, such analytic studies must eventually evolve into or be applied to the forms of cumulative meaningful learning that are characteristic of applied school content-area environments.

Considered in context of the research, theory, and perspectives presented in this chapter, the future of research in content-area reading comprehension in general and in science text comprehension in particular (as a content-area exemplar) are highly promising. However, in contrast to the two research approaches (Approach 1, Approach 3) discussed earlier, a major implication of the research cited in this chapter is that the traditional research approach (Approach 2), which uses narrative reading materials for the development of student reading comprehension proficiency, faces a difficult challenge. Instead of providing students with opportunities for cumulative meaningful learning, traditional reading materials typically involve students in reading stories that involve the arbitrary arrangement of common objects (e.g., people, events, dynamics) that minimize cumulative knowledge development (see Romance, Vitale, & Greene, 2003; Walsh, 2003). Although there is an expanding national emphasis on increasing the use of informational text in elementary settings (e.g., Duke, Bennett-Armistead, & Roberts, 2003; Duke, Martineau, Frank, & Bennett-Armistead, 2003; Duke & Pearson, 2002), the focus of present educational curricular policy on reading skills (vs. content-area comprehension) makes it unlikely that traditional environments for reading instruction will begin to emphasize the cumulative development of meaningful knowledge as a foundation for either general or content-specific reading comprehension proficiency (Vitale, Romance, & Klentschy, 2005, 2006). Such paradigmatic changes in curricular policy properly will require a consensus of research findings that are supportive of effectiveness, which in turn will require a greater emphasis in reading comprehension research on addressing the question of the ecological validity of such studies for enhancing content-area reading comprehension in applied school settings.

Considering reading comprehension as a variant of general comprehension is suggestive of a framework for studying content-area reading comprehension. As we have presented in this chapter, this perspective magnifies the central role of knowledge in developing student reading comprehension proficiency as a form of expertise that is applicable to both domain-specific and general (i.e., transferable to) settings for learning that research seeks to explain. By incorporating the present understanding of the role of knowledge in cumulative, meaningful learning (see Bransford et al., 2000), the application of reading comprehension and reading comprehension strategy research to the problem of content-area reading comprehension in applied school settings has the potential to advance far more quickly.

ACKNOWLEDGMENTS

Preparation of this chapter was supported by Institute of Education Sciences (IES) Project R305G04089 and National Science Foundation/Interagency Education Research Initiative (IERI) Project REC 022835.

REFERENCES

Adams, G. A., & Engelmann, S. (1996). *Research on Direct Instruction: 25 years beyond DISTAR*. Seattle, WA: Educational Achievement Systems.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369–403.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review*, 94, 194–210.

Anderson, J. R. (1992). Automaticity and the ACT theory. *American Journal of Psychology*, *105*, 165–180.

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist*, *48*, 35–44.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*, 335–365.

Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, Cognition*, *20*, 1322–1340.

Armbruster, B. B., & Osborn, J. H. (2001). *Reading instruction and assessment: Understanding IRA standards*. New York: Wiley.

Artzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *Psychological Record*, *47*, 309–320.

Beane, J. A. (1995). Curriculum integration and the disciplines of knowledge. *Phi Delta Kappan*, *76*, 646–622.

Billmeyer, R., & Barton, M. L. (1998). *Teaching reading in content areas*. Alexandria, VA: Association for Supervision and Curriculum Development.

Block, C. C., & Pressley, M. (Eds.). (2002). *Comprehension instruction: Research-based best practices*. New York: Guilford.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn*. Washington, DC: National Academy Press.

Carnine, D. (1991). Curricular interventions for teaching higher order thinking to all students: Introduction to a special series. *Journal of Learning Disabilities*, *24*, 261–269.

Dalton, B., Palincsar, A. S., Defrance, N. L., & Hapgood, S. E. (2006, April). *The role of strategy instruction in supporting students' understanding of graphic intensive science hypertext*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Dick, W., Cary, L., & Cary, J. L. (2004). *The systematic design of instruction* (6th ed.). New York: Allyn & Bacon.

Dillon, T., & Tan, P. L. (1993). *Object-oriented conceptual modeling*. New York: Prentice Hall.

Dougher, M. J., & Markham, M. R. (1994). Stimulus equivalence, functional equivalence and the transfer of function. In S. C. Hays, L. J. Hays, M. Santo, & O. Koichi (Eds.), *Behavior analysis of language and cognition* (pp. 71–90). Reno, NV: Context.

Dreher, M.. J. (2002). Children searching and using information text. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 289–304). New York: Guilford.

Duke, N. K., Bennett-Armistead, V. S., & Roberts, E. M. (2003). Filling the nonfiction void. *American Educator*, *27*(1), 30–35.

Duke, N. K., Martineau, J. A., Frank, K. A., & Bennett-Armistead, V. S. (2003). *3.6 minutes per day: The impact of including more information text in first grade*. Unpublished manuscript, Michigan State University.

Duke, N., & Pearson, P. D. (2002). Effective practices for developing reading comprehension. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 205–242). Newark, DE: International Reading Association.

Ellis, A. K. (2001). *Research on educational innovations*. Larchmont, NY: Eye on Education.

Engelmann, S., & Carnine, D. (1991). *Theory of instruction*. Eugene, OR: Association for Direct Intervention.

Farstrup, A. E., & Samuels, S. J. (Eds.). (2002). *What research has to say about reading instruction*. Newark, DE: International Reading Association.

Florida Department of Education. (2005). *Florida's comprehensive assessment test: 2005 FCAT scores*. Retrieved August 2005 from http://www.firn,edu/doe/sas/fcat/fcpress.html

Gagne, R. M., Wager, W. W., Golas, K., & Keller, J. M. (2004). *Principles of instructional design*. New York: Wadsworth.

Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, *71*, 279–320.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, *39*, 93–104.

Graesser, A. C., Leon, J. A., & Otero, J. C. (2002). Introduction to the psychology of science text comprehension. In J. Otero, J. A. Leon, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 1–15). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Graesser, A. C., McNamara, D. S., Louwerse, M. M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text? In A. P. Sewwt & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82–98). New York: Guilford.

Graesser, A. C., Olde, B., & Klettke, B. (2002). How does the mind construct and represent stories? In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 231–263). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Graesser, A. C., & Wiemer-Hastings, K. (1999). Situation models and concepts in story comprehension. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 77–92). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Guthrie, J. T., Cox, K. E., Anderson, E., Harris, K., Mazzoni, S., & Rach, L. (1998). Principles of integrated instruction for engagement in reading. *Educational Psychology Review*, *10*, 177–199.

Guthrie, J. T., & Ozgungor, S. (2002). Instructional contexts for reading engagement. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 275–288). New York: Guilford.

Guthrie, J. T., Wigfield, A., Barbosa, P., & Perencevich, K. C., Davis, M. H., Scafiddi, N. T., & Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, *96*, 403–423.

Guthrie, J. T., Wigfield, A., & Perencevich, K. C. (2004). *Motivating reading comprehension: Concept-Oriented Reading Instruction*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hirsch, E. D. (1996). *The schools we need. And why we don't have them*. New York: Doubleday.

Hirsch, E. D. (2001). Seeking breadth and depth in the curriculum. *Educational Leadership*, *59*(2), 21–25.

Kearsley, G. P. (Ed.). (1987). *Artificial intelligence and instruction: Applications and methods*. New York: Addison-Wesley.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*, 294–303.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.

Kintsch, W. (2002). On the notions of theme and topic in psychological process models of text comprehension. In M. M. Louwerse & W. van Peer (Eds.), *Thematics: Interdisciplinary studies* (pp. 157–170). Philadelphia: Benjamins.

Kintsch, W. (2004) The construction-integration model of text comprehension and its implications for instruction. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1270–1328). Newark, DE: International Reading Association.

Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Processes*, *39*, 125–128.

Kolodner, J. L. (1993). *Case-based reasoning*. San Mateo, CA: Morgan Kaufmann.

Kolodner, J. L. (1997). Educational implications of analogy: A view from case-based reasoning. *American Psychologist*, *52*, 57–66.

Landauer, T. K. (2002). On the computational basis of learning and cognition: Arguments from LSA. *Psychology of Learning and Motivation*, *41*, 43–84.

Landauer, T. K., & Dumais, S. T. (1996). How come you know so much? From practical problem to new memory theory. In D. Herrmann, C. McEnvoy, C. Hertzog, & M. Johnson (Eds.), *Basic and applied memory research* (Vol. 1, pp. 105–126). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284.

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (in press). *LSA: A road to meaning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Luger, G. F. (2002). *Artificial intelligence: Structures and strategies for complex problem-solving*. Reading, MA: Addison-Wesley.

McNamara, D. S. (2006, April). *iSTART: An automated reading tutor for higher-level strategies*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

McNamara, D. S., de Vega, M., & O'Reilly, T. (in press). Comprehension skill, inference making, and the role of knowledge. In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (1998). *Teaching science for understanding: A human constructivist view*. New York: Academic.

National Center for Educational Statistics. (2000). *NAEP 1999 trends in academic progress: Three decades of student performance* (NCES 2000-469). Washington, DC: Author.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of scientific research literature on reading and its implications for reading instruction*. Jessup, MD: National Institute for Literacy.

Niedelman, M. (1992). Problem solving and transfer. In D. Carnine & E. J. Kameenui (Eds.), *Higher order thinking: Designing curriculum for mainstream students* (pp. 137–156). Austin, TX: Pro-Ed.

North Carolina Department of Public Instruction (2005). *Reports and statistics: Student testing results*. Retrieved August 2005 from http://www.ncpublicschools.org/reportstats/html

O'Donnell, A. M., Dansereau, D. F., & Hall, R. H. (2002). Knowledge maps as scaffolds for cognitive processing. *Educational Psychology Review*, *14*, 71–96.

Ogle, D., & Blachowicz, C. L. Z. (2002). Beyond literature circles: Helping students comprehend informational texts. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 259–274). New York: Guilford.

Pearson, P. D., & Fielding, L. (1995). Comprehension instruction. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Volume 2, pp. 815–860). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Perkins, D. N., & Grotzer, T. A. (1997). Teaching intelligence. *American Psychologist*, *52*, 1125–1133.

Romance, N. R., & Vitale, M. R. (1992). A curriculum strategy that expands time for in-depth elementary science instruction by using science-based reading strategies: Effects of a year-long study in grade 4. *Journal of Research in Science Teaching*, *29*, 545–554.

Romance, N. R., & Vitale, M. R. (2001). Implementing an in-depth expanded science model in elementary schools: Multi-year findings, research issues, and policy implications. *International Journal of Science Education*, *23*, 373–404.

Romance, N. R., & Vitale, M. R. (2005, May). *A knowledge-focused multi-part strategy for enhancing student reading comprehension proficiency in grade 5*. Paper presented at the annual meeting of the International Reading Association, San Antonio, Texas.

Romance, N. R., Vitale, M. R., & Greene, E. O. (2003). *A curriculum content analysis of the Scott Foresman Reading/Language Arts: Grades 3, 4, 5*. NSF/IERI Project Technical Report, Florida Atlantic University, College of Education.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H. C., Wiley, D. E., & Cogan, L. S. (2002). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.

Scott-Foresman. (2002). *Scott-Foresman reading for Florida*. Indianapolis, IN: Author.

Schug, M. C., & Cross, B. (1998). The dark side of curriculum integration. *Social Studies*, *89*(2), 54–57.

Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books.

Sidman, M. (1994). *Stimulus equivalence*. Boston: Author's Cooperative.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of Experimental Analysis of Behavior*, *74*, 127–146.

Slavin, R. E. (1990). On making a difference. *Educational Researcher*, *19*(3), 30–34, 44.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, *31*(7), 15–21.

Snow, C. E. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica, CA: RAND.

Sowa, J. F. (2000). *Knowledge representation: Logical, philosophical, and computational foundations*. New York: Brooks/Cole.

Trabasso, T., & Bouchard, E. (2002). Teaching readers how to comprehend text strategically. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 176–202). New York: Guilford.

Vitale, M. R., & Medland, M. B. (2004). *Knowledge structure development: A developmental framework for children's understanding of knowledge*. Boca Raton, FL: Center for School Development.

Vitale, M. R., & Romance, N. R. (2000). Portfolios in science assessment: A knowledge-based model for classroom practice. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 168–197). San Diego, CA: Academic.

Vitale, M. R., & Romance, N. R. (2006a, September). *Concept mapping as a means for binding knowledge to effective content-area instruction: An interdisciplinary perspective*.

Paper presented at the Second International Conference on Concept Mapping, San Jose, Costa Rica.

Vitale, M. R., & Romance, N. R. (2006b). Research in science education: An interdisciplinary perspective. In J. Rhoton & P. Shane (Eds.), *Teaching science in the 21st century* (pp. 329–351). Arlington, VA: National Science Teachers Association Press.

Vitale, M. R., Romance, N. R., & Klentschy, M. (2005, April). *Enhancing the time allocated to elementary science instruction by linking reading comprehension to science: Implications of a knowledge-based model*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Dallas, Texas.

Vitale, M. R., Romance, N. R., & Klentschy, M. (2006, April). *Improving school reform by changing curriculum policy toward content-area instruction in elementary schools: A research-based model*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Walsh, K. (2003). Lost opportunity. *American Educator*, *27*, 24–27.

Weaver, C. A., & Kintsch, W. (1995). Expository text. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 230–245). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Williams, J. P. (2002). Using the theme scheme to improve story comprehension. In C. C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 126–139). New York: Guilford Press.

Yore, L. (2000). Enhancing science literacy for all students with embedded reading instruction and writing-to-learn activities. *Journal of Deaf Studies and Deaf Education*, *5*, 105–122.

II

Using Assessment to Guide Reading Interventions

5

A Multidimensional Framework to Evaluate Reading Assessment Tools

Joseph P. Magliano and Keith Millis
Northern Illinois University

Yasuhiro Ozuru and Danielle S. McNamara
University of Memphis

The assessment of reading comprehension is a critical part of designing and implementing programs that teach reading strategies. This chapter proposes a multidimensional framework for evaluating the appropriateness of reading comprehension assessment tools. We propose that reading comprehension assessment tools should be evaluated in the light of (1) the reading comprehension processes, products, and activities the assessment tool is designed to observe and measure, (2) the ability levels of the target readers, and (3) the types of texts the tool uses to structure and observe examinee reading performance. The framework we present is then used to analyze and evaluate three methods for assessing reading and comprehension: multiple-choice tests of comprehension; short-answer questions designed to measure examinee understanding of the explicit content or the implied situation of a text; and verbal protocols. Through this analysis, we show that the multidimensional framework will play a valuable role when developing new approaches to assess reading comprehension and the use of reading strategies.

The assessment of reading comprehension is a critical part of designing and implementing programs that teach reading strategies. For example, assessing students' reading comprehension ability and skills before an intervention allows potential weaknesses of an individual reader to be diagnosed. Training can then be adjusted to meet the needs of that reader to maximize the impact of the intervention. Effective interventions should also assess and monitor students' progress in the development or improvement of reading comprehension skills throughout the program. Finally, it is usually necessary to assess the extent to which an intervention is effective in improving comprehension skills. For these reasons, evaluation of reading comprehension assessment tools is considered an important first step in designing and developing reading strategy interventions.

In this chapter, we propose a multidimensional framework for evaluating the appropriateness of reading comprehension assessment tools. Specifically, we claim that the effectiveness of a given reading comprehension assessment tool needs to be evaluated by taking into consideration various relevant factors, such as the assessment purpose, the processes and products of comprehension that the assessment is designed to assess, the target examinees, and the texts used in the assessment. A multidimensional framework can be used not only to evaluate existing assessment tools and methods but also to guide the development of new assessment tools. The framework we present is then used to analyze three methods for assessing reading comprehension: (a) multiple-choice tests of comprehension, (b) short-answer questions designed to measure examinee understanding of the explicit content or the implied situation of a text, and (c) the Reading Skills Assessment Tool (R-SAT). The R-SAT was developed by the first and second authors as a method to assess comprehension skills through analysis of think-aloud protocols produced by readers while reading texts (Gilliam, Magliano, Millis, Levinstein, & Boonthum, in press; Magliano & Millis, 2003; Millis, Magliano, & Todaro, 2006). After presenting an overview of each assessment method, we describe the method and results of a correlation study we conducted to evaluate how strongly measures of selected comprehension skills evidenced in verbal protocols are associated with performance on different types of short answer questions. Through this analysis, we hope to show that the multidimensional framework will play a valuable role when developing new approaches to assess reading comprehension and the use of reading strategies.

DIMENSIONS FOR EVALUATING ASSESSMENT TOOLS

The multidimensional framework of reading comprehension presented here was inspired by the general framework of reading comprehension advocated by Snow (2002), which takes into consideration the reader, texts, and reading activities, all of which are bounded by a sociocultural context. Similarly, we propose that reading comprehension assessment tools (hereafter called *assessment tools*) should be evaluated in light of the reading comprehension processes, products, and activities the assessment tool is designed to observe and measure; the ability levels of the target readers; and the types of texts the tool uses to structure and observe examinee reading performance. With respect to this last dimension, we stress the importance of using a discourse analysis, such as a causal network analysis (Trabasso, van den Broek, & Suh, 1989) to explicate the underlying structure of the texts used in an assessment tool. These analyses can be invaluable for predicting comprehension processes and products that should reflect various levels of comprehension at specific points in a text (e.g., Magliano & Graesser, 1991, Trabasso & Suh, 1993) The importance of these dimensions is determined by one's assessment goals. By this, we refer to the reason why the assessment is being conducted as well as the aspect of comprehension targeted by the assessment. This may seem an obvious consideration, but we contend that assessment goals will be met to the extent that they are explicit and evaluated as to whether the tool meets those goals.

Processes, Products, and Activities of Comprehension

Comprehension arises from a series of cognitive processes and activities, including word decoding, lexical access, syntactic processing, inference generation, reading strategies (e.g., self-explanation), and postreading activities (e.g., summarization, question asking and answering, argumentation). These contribute to a reader's ability to connect the meaning of multiple sentences into a coherently connected mental representation of the overall meaning of text. These processes give rise to multiple levels of mental representations (Balota, Flores d'Arcais, & Rayner, 1990; Kintsch, 1988; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983). Although many of these processes can be conceptualized as occurring sequentially on a temporal continuum (Ferreira & Clifton, 1986; Fodor, 1983), many are likely to occur in parallel (e.g., McClelland & Rumelhart, 1981; Wiley & Rayner, 2001), at least for proficient readers.

It is worth noting that theories of comprehension in discourse psychology over the past 20 years have focused almost exclusively on processes that occur during reading (e.g., Balota et al., 1990) rather than on comprehension processes that continue after reading. However, comprehension may develop further even after one has finished reading a text (Bartlett, 1932). This is important when one considers educational settings in which students are asked to engage in activities that use knowledge gained from reading for purposes such as answering questions and/or writing essays drawing from multiple sources. These postreading activities influence the reader's understanding of what was read and generally improve comprehension by helping the reader to reorganize and synthesize the information (see chap. 19, this volume).

Products of comprehension refers to mental representations resulting from comprehension processes. Theories of discourse processing assume that mental representations of texts contain multiple levels of meaning (Fletcher, 1994; Graesser & Clark, 1985; Graesser, Millis, & Zwaan, 1997; Kintsch, 1988; Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983; Zwaan & Radvanksy, 1998). Readers construct a representation of the explicit text content, which is referred to as the *propositional textbase*. This representation contains a network of propositions that capture the explicit ideas contained in a text. The textbase is incrementally constructed in a network as the text is read. Relationships between the textbase propositions are often established when they share an argument (e.g., Kintsch & van Dijk, 1978). However, the textbase is not always sufficient to establish a coherent representation of a text (Giora, 1985). Instead, coherence emerges with the construction of a situation model (Graesser, Singer, & Trabasso, 1994; Magliano, Zwaan, & Graesser, 1999; van Dijk & Kintsch, 1983; Zwaan & Radvanasky, 1998). Readers generate inferences that are based on their world knowledge, which enables them to establish implied relationships between text constituents. As such, the situation model provides an index of text constituents along a number of dimensions, such as agents and objects, temporality, spatiality, causality, and intentionality (Magliano et al., 1999; Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995; Zwaan & Radvansky, 1998). It is important to note that both the textbase and situation model representation are part of a highly integrated network that reflects the underlying meaning of a text (e.g., Graesser & Clark, 1985). Finally, readers may construct information about the rhetorical structure or agenda of an author (Grasser et al., 1997), but readers may not do so unless they have the explicit goal to do so.

Products and processes can be measured online and during reading or offline and after reading. For example, if a researcher wants to measure online inference processes involved in constructing the situation model while a student reads, then the researcher should choose measures that can be obtained during reading. These would include sentence reading times, response latencies to probes presented during reading, and think-aloud protocols, each of which have been shown to be valid measures of situation model construction (e.g., Magliano & Graesser, 1991).

In many educational contexts, an assessment of students' reading ability is often inferred on the basis of "offline" measures, such as answering multiple-choice questions that are presented after the actual reading. It is important to emphasize that these questions are typically answered after an initial reading of the texts, as opposed to directly assessing processes during reading. The decision processes involved in responding to multiple-choice questions introduce cognitive processing tasks that are not relevant to online text processing in a nontest context (chap. 6, this volume; Gorin, 2005), thus causing complex interactions among the text, questions, and answers. As such, these tests assess processes and products of comprehension and of question answering, which is not exactly the same as reading comprehension (see Graesser & Clark, 1985, for a similar perspective regarding open-ended questions). Well-constructed questions and options can sometimes rule out knowledge use and directly tap what gets constructed online, but such designs of question composition are extremely difficult to engineer.

Alternatively, one may be able to engage in a theoretically motivated discourse analysis of items on existing tests to determine the extent to which they provide an indirect assessment of the products of online reading processes. Later in this chapter, we discuss the taxonomy for evaluating different types of questions that occur in the Gates McGinitie (G-M) and Nelson–Denny (N-D) tests of comprehension. For example, readers may be asked to evaluate the meaning of a word in the context of a sentence or text, which would measure lexical processing. On the other hand, a question may ask the student to evaluate an inference that relies on his or her situation model level of understanding. VanderVeen et al. (chap. 6, this volume) developed a similar taxonomy of questions that occur in the Critical Reading Section of the Scholastic Aptitude Test. They provide strong evidence that their classification of question types on this text can be used to create reader profiles that reflect proficiencies in different aspects of comprehension.

Ability Level of Readers

Reading comprehension is a product of complex interactions between the properties of the text and what readers bring to the reading situation. Proficient readers approach a text with relevant knowledge, word decoding ability, text-based and situation model-based inferencing skills, competency with a variety of reading strategies, metacognitive skills, and so on (Graesser, Millis, & Zwaan, 1997; McNamara & O'Reilly, in press; Oakhill, 1994; Perfetti, 1985, 1994; Snow, 2002). Each of these dimensions has a profound impact on comprehension and may hold implications for the assessment of individuals' reading comprehension ability (Hannon & Daneman, 2001; Oakhill, 1994; Perfetti, 1985, 1994).

Dimensions of an individual's reading ability are likely to vary as a function of literacy education or experience and alter their contribution to overall reading ability. For example, whereas inadequate proficiencies in early and lower level processes (e.g., phonological and lexical processes) are a primary reason why beginning readers struggle (Perfetti, 1985, 1994), there is some evidence that the reading abilities of older children are more closely related to differences in higher level reading skills, such as the ability to make text-based or situation-based inferences, to maintain coherence, to activate higher order knowledge structures, or to monitor and manage comprehension processes (e.g., Oakhill, 1994; chap. 3, this volume; Perfetti, 1985; chap. 6, this volume). As a result, assessing vocabulary knowledge and/or word decoding ability to identify at-risk readers may be particularly appropriate during the early stages of literacy training. However, the same assessment tool is likely to fail to identify at-risk readers among older children. Instead, during later literacy training and secondary education, deficiencies in inference processes and strategic comprehension skills are the major roadblocks for students who are trying to learn new information through reading (Snow, 2002). Thus, the target processes or products of assessment need to be adjusted on the basis of the developmental stage of the target students.

Influence of Text Characteristics

Students read text for different purposes, and reading purposes are closely associated with the text genre. For example, some goals for of reading narrative stories may be to understand the basic sequence of events described, be entertained, and extract some moral or point. On the other hand, the primary purpose of reading expository texts such as science or history texts is to learn or acquire new information about scientific or historical facts about natural/social events. In addition, these two types of texts differ in terms of the novelty of information contained in the text. Thus, the same reader may appear relatively strong or weak depending on the reading situations, which often involve different purposes that are largely associated with the text genres (Best, Rowe, Ozuru, & McNamara, 2005; McNamara, Floyd, Best, & Louwerse, 2004). In order to accurately detect the intra-individual differences in reading comprehension resulting from text/genre effect, it is important to take into consideration the notion that different goals are associated with these different types of texts.

Finally, there is evidence indicating that even within a given genre, text characteristics and individual differences interact in determining the reading comprehension performance of a given individual (McNamara & Kintsch, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996). For example, in the context of scientific texts, McNamara and colleagues have shown that low-knowledge readers comprehend high-coherence texts better than low-coherence texts, whereas the opposite is true for high-knowledge readers. Hence, this line of research indicates the presence of intra-individual differences in reading comprehension performance as the function of matching characteristics of the texts (e.g., cohesion) and the individual's knowledge level. Readers appear to comprehend science texts optimally when reading a text that poses a moderate level of challenge.

Overall, the discussion presented in this section indicates a rather complex picture of reading comprehension assessment. One may get different pictures of comprehension depending on the combination of assessment tools, age groups, genre of the text used in the assessment, and text characteristics within a genre. We propose that assessment tools must be chosen and developed with the target reading situation in mind. Assessment tools should contain texts and activities that are representative of those that students actually encounter in the nontest context that the assessment is designed to measure.

As an example, the Discourse Technology Group at Northern Illinois University was recently asked to evaluate the reading comprehension portion of the Law School Admissions Test (LSAT), which uses a multiple-choice format. Law students often encounter and produce argumentative texts. They must be able to construct coherent textbase and situation model representations for this type of text. They must also be able to reason beyond those texts and determine their relationship to other arguments that may occur in the context of a legal case. A careful analysis of sample LSAT problems revealed a relatively equal number of questions that assessed readers' ability to construct a textbase, to generate appropriate inferences in the context of a situation model, and to reason beyond the texts. As such, the LSAT measures a variety of processes and products of comprehension that a law school student is expected to master during the course of his or her training to comprehend, interpret, and argue based on legal documents.

A final point about the text materials used in assessment is that researchers should understand and be sensitive to the structural features (e.g., causal and rhetorical structures) of the texts included in the assessment tool, because these structural features influence the extent to which readers engage in strategic processing of texts (e.g., Trabasso et al., 1989; chap. 14, this volume). Researchers should engage in some form of discourse analysis that provides an understanding of the features of the texts used in the assessment. These analyses can provide insight into the processes and products that a given text affords.

For example, if one wanted to assess the extent to which readers establish bridging inferences between important text constituents, a causal network analysis (Trabasso et al., 1989) could be administered to determine causal relationships afforded by a text. Such an analysis could be used to identify potential causal inferences that skilled readers should generate while reading a particular text (e.g., Suh & Trabasso, 1993, Trabasso & Suh, 1993). As another example, a propositional network analysis advocated by Kintsch and van Dijk (1978) could also be used to identify the breaks in cohesion that skilled readers should be able to resolve (Graesser, McNamara, Louwerse, & Cai, 2004; McNamara et al., 1996; O'Reilly & McNamara, 2006). These analyses could then provide a basis for constructing test activities and items to assess readers' ability to establish coherence during reading.

Assessment Goals

The complexities suggest that selecting an assessment tool should be guided by the specific goal of the assessment. In this section, we discuss issues related to assessment goals to provide more specific guidelines for the selection process. Comprehension assessment may occur in any situation in which a researcher or educator is interested in understanding psychological processes or products of reading. Assessment may occur for a variety of purposes: evaluating the fluency of online processing of materials, assessing the nature of a memory representation, or determining how effectively a student can apply the knowledge gained from a text to a relevant task (e.g., law school students developing an argument based on legal materials). Assessment occurs in variety of contexts that range from laboratory to educational settings. The same assessment techniques are not appropriate in all settings. For example, in the context of discourse psychology research, the primary goal of the assessment may be to identify the nature of inference processes that occur online during reading (e.g. Graesser et al., 1994; Magliano & Graesser, 1991; McKoon & Ratcliff, 1992). As such, researchers have used a variety of tasks that provide measures of reading behavior, such as sentence reading times, eye movements, probe response methodologies (e.g., lexical decision, word-naming), and verbal protocol methodologies (e.g., thinking aloud). Many of these methodologies could not be readily implemented in educational settings for both practical and institutional reasons. In addition, an emphasis on test-based accountability has resulted in individual state governments mandating the use of standardized assessment tools (Dwyer, 2005). As a result, alternative assessment approaches for evaluating student achievements used by discourse and school psychologists (Deno, 1985; Shinn, 1989) may not be readily adopted.

In the context of this volume, it is important to consider the extent to which assessment tools provide a basis for guiding reading strategy interventions. Reading comprehension assessment in the context of strategy interventions can roughly be classified into two categories based on the goals such assessment is designed to achieve: (a) general classification of readers and (b) diagnosing readers' specific weakness or problems.

The first type of assessment is intended to provide a general classification of readers rather than providing a detailed diagnosis of specific problems in reading comprehension. For example, before an intervention, students are typically assessed to determine whether they are at risk or experiencing reading problems. However, once at-risk students are identified, the second type of assessment needs to be administered to afford a more detailed diagnosis of the locus of students' problems within the reading comprehension process. Students may have difficulty at decoding, lexical access, or higher level comprehension skills, such as inference-making. Detailed diagnosis of the students' problems allows educators to determine the type of intervention that can specifically target the weaknesses or the problems exhibited by the students.

Thus, these two types of assessment goals become highly relevant constraints when selecting an appropriate assessment tool because assessment tools vary in terms of fulfilling the requirement associated with these two goals. Many standardized assessment tools that have been designed to provide a general assessment of a reader's ability to comprehend text appear to be suited for classifying readers into skilled and unskilled readers who require intervention (chap. 6, this volume). The standardized general reading ability tests, such as the G-M and N-D tests, for example, use a multiple-choice format in which readers comprehend a series of short texts and answer multiple-choice questions regarding different aspects of their understanding of the text. Although these tests are not without their shortcomings (Carver, 1992; Farr, Pritchard, & Smitten, 1990; Hanna & Oaster, 1980; Katz, Lautenschlager, Blackburn, & Harris, 1990), they are quite effective for classifying students because they are readily available, cheap to administer, and have been shown to be reliable and valid assessments of general reading skills (Freedle & Kostin, 1994; Glover, Zimmer, & Bruning, 1979; Malak & Hageman, 1985; van den Bergh, 1990).

However, many of these standardized tests are not designed to provide a detailed picture of why less skilled or at-risk students are comprehending texts poorly. Less skilled students' performance on these tests could result from problems or deficits within any phase of the reading comprehension processes. This shortcoming is in part related to the fact that the construct of reading comprehension on which these tests are based is not explicitly informed by a substantive psychological model of reading processes based on research in discourse theory. Consequently, these tools have, thus far, shed little light on specific reading deficits and their remediation. Assessment tools that provide a general assessment of comprehension proficiency will not be sufficient if one's goal is to foster the development of proficiencies lacking in a specific student. Rather, one must diagnose an individual reader's deficits, because there are multiple reasons why a student may struggle to read. To meet this goal, a battery of tests tapping lower and higher level processes would most likely be needed. Nonetheless, it may be possible to make use of constructs in discourse theory to determine which aspects of comprehension are tapped by assessment tools (chap. 6, this volume). In the next section, we describe a taxonomy that was developed to determine the processes and products of comprehension that are measured by the G-M and N-D tests.

Researchers often develop their own assessment tools in addition to these standardized tests. One tool commonly used among discourse researchers is short-answer questions that assess memory for the propositional textbase and situation model representations for a text (e.g., Magliano et al., 2005; McNamara, 2004). For example, the question "What representations are assessed by short-answer questions?" could be used to assess the comprehension of this paragraph. Short-answer questions require examinees to access specific aspects of their memory representation produced while reading. To the extent that accessibility of the specific information based on a given cue (i.e., question stem) is largely a function of the processing performed at the time of reading, memory-based short-answer questions may tap the representation formed at the time of reading the passage more directly than multiple-choice reading comprehension questions. If the questions are presented to the readers after the reading, the questions do not influence the reading process. Short-answer questions require readers to generate the answer themselves on the basis of the question stem, which makes a short-answer question distinct from multiple-choice questions, which can be answered partly on recognition memory, information search in the target passage, and reasoning. It is important to note that short-answer questions also have limitations. As we discuss in the following section, readers can sometimes provide correct answers to short-answer questions using strategies that are not always indicative of their comprehension ability. As a consequence, it may be difficult to identify a specific locus of a comprehension problem or deficit.

In the context of strategy interventions, identifying which aspects of comprehension improve as a function of training is critical. It is particularly important to demonstrate that students adopt the reading strategies addressed in an intervention. However, neither multiple-choice reading comprehension questions nor short-answer questions provide a direct measure of online reading processes and strategies. This is especially important given growing evidence that different populations of readers differentially benefit from training (e.g., Magliano et al., 2005; McNamara, 2004; chap. 16, this volume). As such, if the goal is to assess how reading behaviors changes as a function of an intervention, then alternative measures that are sensitive to reading strategies, such as think-aloud protocols, would need to be adopted.

Think-aloud protocols are well suited to assess of the nature of inferential processing and strategies that students use in an attempt to understand a given text. In the context of the second goal of the assessment, which is to diagnose specific weaknesses and problems that readers face in the temporal continuum of reading, we believe that think-aloud protocol analysis may provide the valuable tool, in addition to other forms of assessment, in determining or designing specific interventions.

EVALUATING ASSESSMENT TOOLS

In this section, we use a multidimensional approach to evaluate three different techniques for assessing comprehension. The first technique involves multiple-choice questions, the second involves open-ended questions, and the third involves think-aloud protocols. In particular, we evaluate these three approaches in terms of the processes and products measured by each approach, the nature of the texts in relation to the target reader, and the goals for implementing or developing the assessment tool. We present results of analyses performed on existing corpora of multiple-choice questions (G-M test and N-D test), short-answer questions, and think-loud protocols, and compare them to assess relationships among these different approaches.

Multiple-Choice Tests of Reading Comprehension

Multiple-choice tests of reading comprehension are arguably the most common of the three approaches. Although it is possible to construct these tests on the basis of discourse theory, many tests are constructed only on the basis of the psychometric properties of the items, and as a result one may be not be appropriate to assess many of the processes and products of comprehension as outlined by theories of discourse comprehension. As such, researchers and educators cannot assume that such tests adequately measure dimensions of interest in reading comprehension. In an effort to rectify this situation, we have developed a taxonomy that classifies assessment questions in terms of the nature of comprehension that the questions assess (i.e., type of processes and resulting representation, such as textbase, situation model). We used the taxonomy to classify the questions in two commonly used assessments of general reading comprehension, the G-M and N-D reading comprehension tests. We used Form T of the G-M and Form F of the N-D, which are both used to assess late adolescent readers' comprehension abilities (Grade 12 and college freshmen).

In an analysis of the N-D and G-M tests, we identified at least three general classes of questions. The question types differ on the processes and products of comprehension that they address. Example questions from the N-D and G-M for each type are shown in Table 5.1. The first class of questions is *local textbase questions*. The two processes involved in answering these questions are (a) searching and locating the explicit text content in a particular sentence and (b) verifying which answer most closely matches the text content. These questions require minimal if any inferential processes. The product of comprehension most closely associated with this question is the textbase. We consider these questions local because students have to consider only one or two adjacent sentences to answer the question.

The second class of questions is *global-textbase questions*. This question type differs from the local-textbase questions on the grain size of text that the reader is asked to consider. The answers to local-textbase questions are usually found in one sentence. With global-textbase questions, however, the reader is asked to determine whether a phrase or word reflects the thematic meaning of a segment of text longer than one sentence (e.g., several sentences, paragraph, or entire texts). In this sense, the potential answers provide paraphrases or summary statements of the theme that are reflected in multiple propositions in the textbase representation. However, these questions may also require reference to the situation model to the extent that the reader may generate thematic inferences

TABLE 5.1

Example Questions for Local, Global, and Inference Questions

Type	*Stem*	*Answer Options*
Local	The audience size mentioned was	a. five thousand b. eight thousand c. twelve thousand d. sixteen thousand e. twenty thousand
Global	This passage is mainly about how one kind of spider	a. excavates a tunnel b. traps its food c. fools it's enemies d. builds with silk
Inference	When Elizabeth's parents were watching the show, they were	a. impressed b. nervous c. ashamed d. proud of themselves
Inference	The passage suggests that the teacher would have thought that today's cars are	a. easier to drive than Model T's b. more fun than Model T's c. more like locomotives than Model T's d. more complicated to fix than Model T's

or generalizations of the text segments. In terms of processes, the reader must search and locate the appropriate segments of the texts, construct a summarization of that segment, and then assess which answer option best matches that summarization. The example global question in Table 5.1 was classified as such because the reader must consider the entire text to identify the appropriate answer.

The third class of questions is *inference questions*. These require processes associated with generating inferences from the texts. The nature of the inference depends on the question. The inferences could be bridging, explanatory, predictive, or elaborative. Some of these questions may assess inferences that are considered to be normally generated while reading the texts (e.g., Graesser et al., 1994). As such, the comprehension product associated with these questions corresponds to the situation model (see Zwaan & Radvansky, 1998, for a review) that proficient readers are considered to form in the normal course of reading. The first example inference question in Table 5.1 was classified as such because it involves inferring the affective response of characters in a short narrative. Other inference questions may require readers to reason beyond the texts and generate inferences that are specific to the content of the questions. The second inference question in Table 5.1 is an example of this type of question because it requires the reader to reason beyond the text, because it never discusses modern automobiles. It is important to note that not all inference questions in these tests reflect inferences that readers would normally construct when reading the texts. This is particularly the case for the second class of inference questions. Inference questions that tap situation model inferences would need to be carefully constructed in light of discourse comprehension theory. One may consider inference questions to be the most difficult because they require the reader to reason beyond the explicit text. However, the difficulty of these questions will be determined by the extent to which the text supports the inference and similarity of the alternative answer to the correct inference.

We analyzed the frequency of these different types of questions for the G-M and N-D tests. With respect to the G-M, 56% (*N* = 27), 13% (*N* = 6), and 31% (*N* = 15) of the questions were local textbase, global textbase, and inference questions, respectively. With respect to the N-D, 58% (*N* = 22), 8% (*N* = 3), and 33% (*N* = 13) were local textbase, global textbase, and inference questions, respectively. It is clear that both tests primarily measure processes associated with verifying the textbase. However, it is important to note that both tests contain a substantial percentage of questions that assess inferences associated with the situation model.

We also analyzed the passages used in the tests. As discussed earlier, text genre and text structure influence the nature of the examinee–text interaction and should be accounted for by assessment tools. The G-M contains a collection of 11 short texts, 5 of which were narrative texts and 6 of which were expository texts. All but of 2 the texts contained only one paragraph with a mean length of 122 words. We calculated reading grade level with the Flesch-Kincaid to determine whether the texts are appropriate for late adolescent and early adult readers. The grade level for the texts ranged from 7.3 to 12.0, and the mean grade level was 10.4. With respect to the N-D test, there were eight texts, all of which were expositions. All texts but one had two to three paragraphs with a mean length of 266 words. The Flesch-Kincaid grade level ranged from 6.6 to 12.0, with a mean of 10.0. The texts seemed appropriate for late high school and college students. In fact, both tests contained outlier texts that lowered the average grade level. Because the G-M contained both narrative and expository texts, it is more representative of the texts that students might encounter in academic reading situations in high school.

With respect to the assessment goals associated with these two tests, they are primarily used to identify skilled and less skilled readers (e.g., Magliano & Millis, 2003). However, the analyses of the different question types may allow one to identify proficiencies in constructing local textbase, global textbase, and situation model representations. We assessed the possibility that performance on the different question categories is correlated with short-answer questions that require access of the textbase or situation model representations.

Short-Answer Tests of Reading Comprehension

Another approach to assessing comprehension is using short-answer questions (e.g., Magliano et al., 2005; McNamara & Kintsch, 1996). In a typical short-answer question, readers read text and then answer the questions without looking back at the text. Answers can range from a single word to several clauses. Answering questions in this type of assessment requires a reader to access the memory representation for a text and retrieve and produce the relevant information.

The questions are typically designed to provide an assessment of the quality of either the explicit textbase or the situation model. The Appendix contains an example of a text and questions. The textbase questions require the reader to retrieve information that can be found in a single sentence of the text. For example, Question 10 for the text on the Franco Dictatorship is "In what year did the Franco dictatorship end?" Its answer can be found in the last sentence. The majority of the situation model questions assess the extent to which readers have inferred causal relationships between text events. For example, consider Question 7, "What were the causes of the great period of economic stagnation that followed World War II?" This question can be answered by the content of several sentences in the fourth paragraph of the text. The extent to which readers can answer these questions should be related to the extent to which they generated the causal inferences. It is important to note, however, that these questions measure memory for both explicit content and inferred relationships.

Of course, there are some limitations of using short-answer questions. First, readers might have known the answer before reading the text. In this case, the test question would be measuring not their comprehension ability but rather their prior knowledge. Second, one cannot completely rule out the possibility that comprehension occurs as the readers search their memory representations as they try to answer the question. A reader's memory representations might be sufficient for him or her to successfully use reasoning and guessing strategies. The last limitation concerns the scoring process. Scoring the answer requires identification of the ideal answers that might contain several parts, as is the case with situation model questions. Each participant's answer must be classified with respect to the percentage of the important parts present within the answer. This type of practical limitation becomes a large factor when one needs to assess general reading skill for a large number of students in a short period of time before an intervention.

Think-Aloud/Verbal Protocols

The first and second authors are developing an alternative to multiple-choice and short-answer questions for assessing reading comprehension. The assessment tool is called the *Reading Strategy and Assessment Tool* (R-SAT, Gilliam et al., in press), and it uses verbal protocols. The R-SAT is designed to measure comprehension strategies associated with different standards of coherence. *Standards of coherence* refers to a reader's criteria or general sense of the importance of forming a coherent representation, especially of how different parts of a text are related to one another (van den Broek, Risden, & Husebye-Hartman, 1995). As stated earlier, deep comprehension arises with the construction of coherent textbase and situation model representations. However, van den Broek et al. (1995) argued that readers differ in the extent to which they have a drive to achieve coherent representations. Some readers will accept disjointed representations of the explicit propositions contained in texts, whereas other readers attempt to construct a representation that contains coherent relationships between those propositions.

Our past research has consistently demonstrated that think-aloud protocols can reveal a reader's standards of coherence. Magliano and Millis (2003; Millis et al., 2006) had participants think aloud when reading selected sentences in simple narratives (i.e., Chinese folk tales). Less skilled readers, as identified by the N-D test, talked more about the target sentence than did skilled readers, who talked more about how those sentences were related to the prior discourse context. Thus, skilled readers' protocols show their effort to maintain larger or more global coherence, whereas less skilled readers' protocols show their tendency to focus on target sentences in isolation. Magliano and Millis used latent semantic analysis (LSA; Landauer & Dumais, 1997) to provide a quantitative assessment of these strategies. LSA provides a measure of overlap between any two units of language by computing cosines between their vector representations in a high dimensional semantic space. These cosines typically range from 0 to 1.0 and represent the degree of conceptual relations between the linguistic elements. Magliano and Millis computed cosines between the think-aloud protocols of a target sentence and two semantic benchmarks (i.e., the target sentences and causally important prior text sentences) and analyzed participants' recall performance for different texts that were read silently in terms of these two types of LSA cosines. The analysis indicated that recall performance decreased as the function of the LSA cosine between the think-aloud protocol and the target sentences that was just read (large conceptual overlap between the targets sentence and the protocol), whereas the recall increased as the function of the LSA cosine between the protocol and the causally important sentences (large overlap between causally important sentence and the protocol).

Thus, two lines of results (i.e., relation between protocol and reading skill, and relation between protocol and subsequent recall) converge, establishing that the R-SAT is capable of revealing readers' online processing involved in the maintenance of text coherence.

In the R-SAT, readers are provided with questions about the content of the sentence (e.g., "Why did the battle fail?") and questions to facilitate thinking aloud ("What are you thinking right now?") while reading. The questions appear immediately after the presentation of preselected sentences. These sentences were preselected on the basis of the presence of strong causal connections to prior portion of the text. Correct and complete answers to both types of questions require that the reader generate the appropriate causal bridging inferences at that point in the text. With the R-SAT, readers produce their answers and think-aloud protocols by typing them into the computer. LSA and word-matching algorithms are then used to assess the completeness of the answers by comparing them to ideal answers and assess the extent to which the think-aloud protocols conceptually overlap with the current target sentence and causally important target sentences.

Overall, the R-SAT provides important information to assess reading comprehension in terms of both process and product of comprehension. With respect to processes of comprehension, the R-SAT assesses the extent to which readers are generating causal bridging inferences based on think-aloud protocols produced in response to a think-aloud question. With respect to products of comprehension, the R-SAT assesses the quality of the textbase and situation model representations based on the readers' answer to the think-aloud question.

Discourse analyses of the texts used in the R-SAT are a central component to its development. Specifically, a causal network analysis (Trabasso et al., 1989) was used to identify local, distal direct, and distal indirect causal consequences in the prior discourse context. This analysis provides a basis for constructing the important prior text information that is compared to the think-aloud protocols to assess the extent to which readers are generated bridging inferences.

Finally, to be consistent with the multidimensional framework, it is important to assess the intended goal for the R-SAT. We believe that the R-SAT is appropriate for providing a general assessment of reading skill to the extent that reading skill is influenced by a reader's standard of coherence. Another goal of the R-SAT is that it will be integrated into iSTART, a reading intervention designed to increase standards of coherence by teaching students how to self-explain as they read (see chap. 16, this volume). We foresee that iSTART training will be tailored to individual readers' needs based on assessments provided by the R-SAT. For example, students who show low standards of coherence as indicated by the R-SAT will be taught the rudiments of self-explanation by means of iSTART. However, the training for students who show relatively high standards before training could focus on fine tuning self-explanation, such as determining the appropriate sentences to self-explain. The general approach used in the R-SAT has also been used to assess ongoing progress during practice sessions in iSTART (e.g., Magliano, Wiemer-Hastings, Millis, Muñoz, & McNamara, 2002). Specifically, the self-explanations produced during practice are compared to three things: (a) the sentence just read, (b) prior discourse, and (c) concepts related to but not present in the discourse context. Success in training can be assessed by measuring the extent to which the protocols overlap with these benchmarks. For example, students whose thoughts overlap with the sentence just read, and not with either the prior discourse or concepts from world knowledge, are most likely not explaining the text but rather only paraphrasing or repeating the text. As a final goal of the R-SAT, we have used this general approach to assess whether readers change their reading behaviors after training (Magliano et al., 2005). We have found that, in general, readers' thoughts overlap more with the discourse context after training than before training. More specifically, self-explanation protocols after training contain more concepts from the current sentence and prior discourse than before training. This suggests that, after training, readers were explaining how the current sentence fit into the larger discourse context. The R-SAT could provide a basis for making more detailed evaluations regarding changes in reading strategies as a function of iSTART.

ASSESSING CONVERGENCE BETWEEN THE DIFFERENT ASSESSMENT TOOLS

In this section, we present two sets of analyses designed to evaluate relations between the different assessment approaches described in this chapter. The aim of these analyses is to determine the extent to which these assessment tools measure the intended processes and products of comprehension. The first analysis was conducted to determine convergences between the G-M and short-answer tests, whereas the second was designed to assess convergence among the R-SAT, short-answer tests, and the N-D test.

Convergence Between G-M and Short-Answer Performance

As described above, the G-M test comprises three general types of questions that address different aspects of comprehension. Local questions tap explicit ideas in the textbase representation; global questions tap thematic ideas in the textbase, and perhaps situation model representations; and inference questions address the quality of situation model level representation. We had 223 college freshmen and sophomores take the G-M test of reading comprehension. They also read two texts and answered 10 short-answer questions for each immediately after reading the texts. Five of the 10 questions were textbase questions, and the other 5 were situation model questions. The Appendix contains one of the texts and questions that were used in the study. The textbase question could be answered via the content of a single sentence, whereas the situation model questions require readers to infer causal relationships across text sentences.

We assessed performance on the local, global, and inference questions on the G-M test. On the basis of theories of discourse comprehension that assume that the textbase reflects a more shallow level of understanding than the referential situation model, one would expect that performance on the local questions to be the best, followed by global questions and then inference questions. With respect to these latter two categories, one would expect that the inference questions would be harder to answer if these questions required one to consider the deeper, underlying meaning or implications of the discourse. However, the mean percentages of questions answered correctly was 66%, 60%, and 76%, for local, global, and inference questions, respectively. A one-way analysis of variance revealed that these means were significantly different from one another, $F(2, 610) = 94.92$, $p < .05$. A post hoc analysis (Tukey) revealed that inference questions were answered more accurately than local questions, which were in turn answered more accurately than global questions. These results suggest that, contrary to common assumptions, these questions may not be measuring deep comprehension. It is important to note that this conclusion is limited to the G-M. For example, our discourse analyses of the inference questions on the LSAT suggest that these questions do require one to have a deep comprehension of the texts on that test, although we do not yet have data to empirically support this claim.

The short-answer questions were scored on the basis of the proportion of the key ideas present in each answer. We then added up the proportion scores for each participant, yielding the total number of questions answered completely out of a total possible score of 5. Finally, we calculated the average textbase and situation model questions answered correctly over the two texts. As one would expect, textbase questions ($M = 2.01$, $SD = 0.99$) were answered more accurately than situation model questions ($M = 1.51$, $SD = 0.88$), $t(1, 223) = 9.75$, $p < .05$.

One interesting question is how processes used to respond to short-answer questions (i.e., textbase and situation model questions) and multiple-choice questions are related. One reasonable expectation would be that answering local and global questions in the G-M test should be more closely related to performance on textbase questions, because these questions require an intact textbase for them to be answered correctly. It may also be the case that the local questions, as opposed to global questions, on the G-M test better reflect the skills necessary to answer the textbase questions rather than the global questions, because answering the local questions require test-takers to identify specific propositions in a text and evaluate which answer best reflects that proposition. This process is conceptually similar to processes required for answering textbase short-answer questions. The other important prediction is that performance in answering the G-M inference questions should account for most of the variance in performance answering the situation model short-answer questions, because inference questions are supposed to tap readers' ability to construct a situation model by generating inferences. Bivariate correlations between the short-answer and multiple-choice questions are presented in Table 5.2.

To pursue these questions concerning the relations between these two types of assessment tools (i.e., multiple-choice questions in the G-M test and short-answer questions), multiple regression analyses were conducted in which the predictor variables for each regression analysis were the percentages of local, global, and inference questions in G-M test answered correctly. We performed two multiple regressions, one using the textbase short-answer question performance and the other using the situation model short-answer question performance as the criterion variables. The beta weights and R^2s can be found in Table 5.3. The regression analysis on textbase short answer questions accounted for a significant 31% of the variance, $F(2, 219) = 33.40$, $p < .05$. Performance on local, $t(222) = 4.27$, $p < .01$, and inference questions, $t(222) = 2.23$, $p < .05$, were significant predictors of performance on textbase questions. As one would expect, performance on local questions appeared to be the strongest predictor of performance on the textbase short-answer questions. This is in line with our prediction that answering textbase questions involves accessing a specific proposition, essentially similar to the process underlying the local question answering in G-M test. It is quite surprising that performance on the global questions was not a significant predictor of performance on the textbase questions. On the other hand, inference questions were indicative of performance on the textbase short-answer questions. We discuss the implications of these findings shortly.

The regression analysis on situation model questions accounted for a significant 42% of the variance, $F(2, 219) = 53.166$, $p < .05$. Performance on local, $t(222) = 3.77$, $p < .05$; global, $t(222) = 4.64$, $p < .05$; and inference questions, $t(222) = 2.97$, $p < .05$, were significant predictors of performance on situation model questions. The beta weight suggests that the different G-M questions were comparable predictors of performance on these questions. It is interesting

TABLE 5.2

Correlation Matrix for the Textbase, Situation Model, Local, Global, and Inference Questions

	TB	*SM*	*L*	*G*
Texbase(TB)				
Situation Model (SM)	0.68			
Local (L)	0.54	0.57		
Global (G)	0.31	0.49	0.44	
Inference	0.50	0.55	0.73	0.42

TABLE 5.3

Regression Beta Weights Predicting Short Answer Comprehension Performance from the Local, Global, and Inference Questions on the G-M test of Reading Comprehension

	Short Answer Question Type	
G-M PREDICTORS	*Textbase*	*Situation Model*
Local	.36**	.29**
Global	.07	.27**
Inference	.21*	.22**
R^2	.31**	.42**

Note. *$p < .05$; **$p < .01$.

to note that global questions were significant predictors of short-answer situation model questions but not textbase questions. This suggests that there is indeed some overlap in the processes required to construct and evaluate a thematic representation of a discourse and constructing and accessing the referential situation model. Moreover, the situation model questions required readers to access relationships between distal texts sentences that are strongly implied by the texts. This is an aspect of comprehension that likely involves both constructing a coherent textbase and a referential situation model.

Relations Among R-SAT, Short-Answer, and N-D Performance. Millis et al. (2006) assessed the extent to which the R-SAT is related to constructing textbase and situation model representations. Specifically, they were interested in the extent to which overlap between the protocols and the different semantic benchmarks is related to answering short-answer questions that assess the textbase or situation model representation. Three semantic benchmarks were used in the study: (a) current sentence, (b) local sentences (immediately prior sentences), and (c) distal causal sentences. As discussed earlier, overlap with the current sentence and local sentences should reflect processes involved in constructing a local textbase representation for explicit ideas, whereas overlap with the distal causal sentences should be related to processes associated with constructing a coherent global textbase and building a situation model representation. They also assessed the relationship between the R-SAT and the performance on N-D tests; that is, they examined which of the three types of R-SAT scores, as represented by the protocol's conceptual overlap with current sentence, immediate prior sentence, or distal causal sentence, predicted performance on the N-D test. Given that the majority of questions in the N-D test tap local processes, one would expect that overlap with the benchmarks associated with local processing (current and local sentences) should be most indicative of performance on this test.

Participants thought aloud after every sentence while reading science texts that had an average grade level of 6.8. They calculated LSA cosines between think-aloud protocols for a given sentence and the three benchmarks and computed an average cosine for each benchmark for each participant. Participants answered textbase and situation model short-answer questions for the text and took the N-D test of reading comprehension. The average cosines for the three benchmarks were used as predictor variables in three regression analyses predicting performance on the short-answer textbase questions, short-answer situation model questions, and the N-D multiple-choice reading ability questions. The bivariate correlations between the variables are presented in Table 5.4, and the resulting standardized coefficients and R^2s are presented in Table 5.5. Each equation was significant, indicating that the LSA cosines predict performance on these three types of reading comprehension questions. As in Magliano and Millis (2003), the cosines for the current sentence were negatively correlated with the question-answering performance, whereas the cosines for prior causal antecedents were positively correlated with the question-answering performance. This general pattern indicates that comprehension of the text was best when the verbal protocols contained information related to the causal antecedents as opposed to when the protocol primarily contained information related to the current sentence.

The pattern of relations between the protocol and local or distal benchmarks is largely consistent with the expectation based on the theory of discourse processing. Specifically, overlap with the current sentence was negatively correlated with the two outcome measures that most closely reflect understanding of ideas in the explicit textbase, namely textbase questions and the N-D. Although this relationship is consistent with Magliano and Millis (2003), this may seem counterintuitive, because one would expect that the explicit ideas would be better represented when readers talked about them while thinking aloud. However, this relation may exist because readers who tend to talk more about the current sentence may be doing so at the expense of constructing important bridging inferences to the prior discourse context. As a

TABLE 5.4

Correlation Matrix for the Textbase Questions, Situation Model Questions, Current Sentence, Benchmark, Local Causal Benchmark, and Distal Causal Benchmark

	ND	*TB*	*SM*	*CS*	*L*
Neslon-Denny (ND)					
Textbase (TB)	0.28				
Situation Model (SM)	0.24	0.48			
Current Sentence (CS)	–0.33	0.00	0.11		
Local (L)	–0.02	0.21	0.17	0.67	
Distal	0.01	0.36	0.36	0.50	0.65

TABLE 5.5

Regression Beta Weights Predicting Measures of Comprehension (Textbase and Situation Model Short Answer, and Nelson Denny) from LSA Variables

	Measures of Comprehension		
LSA Predictors	*Textbase*	*Situation Model*	*Nelson Denny*
Current sentence	–.36**	–.06	–.61**
Local cause	.09	.01	.29**
Distal cause	.45**	.41**	.11
R^2	.22**	.17**	.19**

Note. * $p < .05$; ** $p < .01$.

consequence, this may result in the formation of isolated or fragmentary representations of the text. Because accessing specific textbase content in memory is partly a function of the connections between the textbase content and other related information in memory, failing to draw these bridging inferences is likely to cause retrieval problems at the time of answering the short-answer textbase questions. Indeed, overlap with the distal causal sentences was related not only to performance on the situation model questions, which tapped the readers' understanding of causal relationships between text constituents, but also to textbase question answering performance. These results indicate that the construction of globally coherent representations with the support of causal bridging inference is critical for retrieving a variety of textual information. The results also indicate, as expected, that overlap between the protocol and the local sentence is positively correlated with overall N-D test performance. This finding converges with our earlier observation that majority of questions on the N-D test (58%) are local questions. It is, however, important to note that the bivariate correlations between the benchmarks are high and could be causing suppression effects. We are currently exploring ways to minimize these correlations.

Finally, we also examined the unique variance accounted for by the N-D and the LSA analysis of the verbal protocols in predicting the textbase and situation model scores. The LSA cosines predicted 15% and 16% of the variance of the textbase and situation model scores, respectively, whereas the N-D predicted only 3% and 6% percent. We believe that these differences indicate that the R-SAT is more sensitive than the N-D to the processes necessary to construct a coherent representation of a text, which primarily consist of questions that address a reader's ability to locate and verify explicit text content. Of course, one should note that the LSA values were computed from the same texts on which the short answers were based and therefore they do not constitute the strongest case for this claim.

CONCLUSION

In this chapter, we presented a multidimensional framework to evaluate assessment tools that is based on Snow's (2002) general framework of reading comprehension. This framework involves an assessment of the products and processes measured by a given tool, the intended reader, text characteristics, and the overarching assessment goals. There are many tools that can be used to assess comprehension. However, not all tools will be sufficient to meet one's assessment goals. For example, if one wants to directly assess a reader's ability to engage in deep reasoning about a text, then an assessment tool that primarily taps processes associated with constructing a textbase representation would not be appropriate. Our goal was to provide a framework for evaluating the utility of an assessment tool given one's assessment goals.

We have illustrated the utility of this framework by evaluating existing assessment tools: multiple-choice reading comprehension questions (e.g., the G-M and N-D tests); short-answer reading questions; and the R-SAT, which is based on think-aloud protocols. With respect to the multiple-choice reading comprehension questions, our corpus analysis of G-M and N-D tests of comprehension using a classification scheme based on the theory of discourse processing indicated that the tests contain questions that assess readers' proficiencies in constructing the local textbase, global macrostructure, and inferences. One concern with respect to the G-M was that participants answered inference questions more accurately than local or global questions. If these questions did indeed tap inferences associated with deep comprehension, then one would have expected these questions to be the most difficult. Indeed, it appears that the global questions were the most challenging. It is important to note that the G-M test did account for an impressive amount of variance in performance on short-answer questions that addressed the textbase and underlying situation model.

It remains to be seen as to whether our discourse analysis of the questions in the G-M could be used to diagnose specific comprehension problems. VanderVeen et al. (chap. 6, this volume) provide an overview of the kind of research necessary to pursue this endeavor with respect the Critical Reading Section of the Scholastic Aptitude Test. Although these researchers did not explicitly use the multidimensional framework advocated in this chapter, their approach is consistent with this framework.

We are considerably more skeptical regarding the N-D test of reading comprehension. Millis et al. (2006) assessed the extent to which found that the N-D test accounted for relatively little variance in performance on short-answer questions that involve both textbase and situation model representations. Furthermore, the extent to which readers established distal bridging inferences was not predictive of performance on the N-D. On the basis of these findings and the analysis of the type of questions contained in the N-D, we are inclined to conclude that the N-D test may not be an ideal assessment tool if one wants to assess a reader's ability to construct a representation that reflects deep meaning. Indeed, to find statistically significant differences between skilled and less skilled readers on various comprehension tasks, we have either used a quartile split (e.g., Magliano & Millis, 2003) or greatly shortened the length of allotted testing time from that recommended by the test publishers (e.g., Magliano et al., 2005; McNamara & McDaniel, 2004).

On the other hand, the results reported here from Millis et al. (2005) suggest that the reading strategies revealed in think-aloud protocols are indicative of inferential processes and products underlying deep-level comprehension, at least to the extent that this measure correlates with one's ability to respond to short-answer questions in a systematic and theoretically predictable way. In general, we find that when readers primarily paraphrase the sentence (evidenced by a large overlap between protocol and current sentence), they tend to have more difficulty answering questions that require access to a textbase representation. The findings also indicated that readers who mention information about the prior text when thinking aloud tend to perform better not only on questions that tap the situation model representation but also on those that tap the textbase representation. We believe that this pattern of results supports the conclusion that individual differences in strategies produced during thinking aloud reflect readers' standards of coherence (Magliano & Millis, 2003). Readers who tend to talk about how the current sentence is related to the prior texts have a higher standard of coherence than readers who tend to only talk about the current sentence; as a consequence, they can construct a more globally coherent representation of the text content that supports access to a variety of information contained in or implicated by the text.

Our evaluation of assessment tools based on the multidimensional framework suggest that the R-SAT (Gilliam et al., in press) is quite a useful tool for assessing readers' ability to engage in effective processing that results in coherent representation of the text content. By combining the R-SAT with the analysis of text properties (e.g., causal network, referential cohesion, word-level analysis), it is possible to obtain detailed pictures of the strengths and weaknesses of individual student's reading comprehension processes: Whereas some students fail to draw inferences when texts do not provide sufficient cues on causal structure, other students may fail when the target sentences include an unfamiliar word or have a complex syntactic structure. For these reasons, the R-SAT shows promise as an approach for assessing reading comprehension, in particular when detailed assessments of the readers are needed to design an intervention or to assess the effect of a particular reading intervention program.

Certainly, the current form of the R-SAT is still in its infancy stage and requires further development. Several plans to improve the R-SAT are in order. First, as seen in Table 5.5, current R-SAT predicts approximately 20 % of the variance (R^2s = .17 and .22) of the textbase and situation model questions answering performance, respectively. This performance is much poorer compared to the G-M test's ability to predict the short-answer question performance (R^2s = .31 and .42) for textbase and situation model questions, which we obtained in a separate study relating the G-M test and short-answer question-answering performance. To improve the performance of the R-SAT, we are currently adopting an approach of identifying specific sentences that are predictive of reading skills and strategies and using only those sentences in the R-SAT (Gilliam et al., in press). We are also exploring the extent to which we can identify specific comprehension strategies used by skilled and less skilled readers, which one could not readily do with traditional multiple-choice tests. It is our hope that we can use this information as a basis for guiding remediation.

The R-SAT has been tested only with adult readers so far. We have tested it with texts well below the reading level of these readers (e.g., Magliano & Millis, 2003) and with more difficult scientific texts (Millis et al., 2006), and obtained similar results. Although the R-SAT could be implemented with younger readers, there is one limitation: Students must be reasonably proficient at typing their thoughts; typing could interfere with students' reading and think-aloud processes if they do not have proficient typing skill. If a version were to be developed for younger readers, it would likely have to be based on orally produced responses.

In the first half of this chapter, we outlined the multidimensional framework for evaluating reading comprehension assessment tools by drawing on recent developments in cognitive psychology and discourse-processing research. Furthermore, we evaluated three types of the existing reading comprehension assessment tools using the multidimensional framework. In the second half of the chapter, we reported recent empirical studies on the relations among three types of the assessment tools: (a) multiple-choice reading comprehension questions, (b) short-answer questions, (c) and the R-SAT, which is based on think-aloud protocols. Overall, empirical findings confirm the validity of theoretical inquiry of the strength and weakness of the three types of assessment tool within the multidimensional framework. Although more research is required to refine the framework, this framework appears to be useful for evaluating and improving tools for reading comprehension assessment.

REFERENCES

Balota, D. A., Flores d' Arcais, G. B., & Rayner, K. (1990). *Comprehension processes in reading*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bartlett, F. C. (1932). *Remembering*. Cambridge, England: Cambridge University Press.

Best, R. M., Rowe, M. P., Ozuru, Y., & McNamara, D. S. (2005). Deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders*, *25*, 65–83.

Carver, R. P. (1992). What do standardized tests of reading comprehension measure in terms of efficiency, accuracy, and rate? *Reading Research Quarterly*, *27*, 346–359.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*, 219–232.

Deno, S. L. (1987). Formative evaluation of individual programs: A new role for school psychologists. *School Psychology Review*, *15*, 358–374.

Dwyer, C. A. (Ed.). (2005). *Measurement and research in the accountability era*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement*, *27*, 209–226.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory & Language*, *25*, 348–368.

Fletcher, C. R. (1994). Levels of representation in memory for discourse. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 589–607). San Diego, CA: Academic.

Fodor, J. A. (1983). *Modularity of the mind*. Cambridge, MA: MIT Press.

Freedle R., & Kostin, I. (1994). Can multiple-choice reading tests be construct valid? *Psychological Sciences*, *5*, 107–110.

Giora, R. (1985). Notes towards a theory of text coherence. *Poetics Today*, *6*, 699–715.

Gilliam, S., Magliano, J. P., Millis, K. K., Levinstein, I., & Boonthum, C. (in press). Assessing the format of the presentation of text in developing a Reading Strategy Assessment Tool (R-SAT). *Behavior Research Methods, Instruments, & Computers*.

Glover, J. A., Zimmer, J. W., & Bruning, R. H. (1979). Utility of the Nelson–Denny as a predictor of structure and themanticity in memory for prose. *Psychological Reports*, *45*, 44–46.

Gorin, J. S. (2005) Manipulating processing difficulty of reading comprehension questions: The feasibility of verbal item generation. *Journal of Educational Measurement*, *42*, 351–373.

Graesser, A. C., & Clark, L. F. (1985). *Structures and procedures of implicit knowledge*. Norwood, NJ: Ablex.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments & Computers, 36*, 193–202.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*, 163–189.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371–395.

Hanna, G. S., & Oaster, T. R. (1980). Studies of the seriousness of three threats to passage dependence. *Educational and Psychological Measurement*, *40*, 405–411.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in component processes of reading comprehension. *Journal of Educational Psychology*, *93*, 103–128.

Katz, S., Lautenschlager, G., Blackburn, A., & Harris, F. (1990). Answering reading comprehension items without passages on the SAT. *Psychological Sciences*, *1*, 122–127.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W., & van Dijk, T. A. (1978). Towards a model of text comprehension. *Psychological Review*, *85*, 363–394.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latend semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Magliano, J. P., & Graesser, A. C. (1991). A three-pronged method for studying inference generation in literary text. *Poetics*, *20*, 193–232.

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure. *Cognition and Instruction*, *3*, 251–283.

Magliano, J. P., Todaro, S., Millis, K. K., Wiemer-Hastings, K., Kim, H. J., & McNamara, D. S. (2005). Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research*, *32*, 185–208.

Magliano, J. P., Wiemer-Hastings, K., Millis, K. K., Muñoz, B. D., & McNamara, D. (2002). Using latent semantic analysis to assess reader strategies. *Behavior Research Methods, Instruments, and Computers*, *34,* 181–188.

Magliano, J. P., Zwaan, R. A., & Graesser, A. C. (1999). The role of situational continuity in narrative understanding. In S. R. Goldman & H. van Oostendorp (Eds.), *The construction of mental representation during reading* (pp. 219–245). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Malak, J., & Hegeman, J. N. (1985). Using verbal SAT scores to predict Nelson–Denny scores for reading placement. *Journal of Reading*, *28*, 301–304.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440–466.

McNamara, D. S. (2004). SERT: Self-Explanation Reading Training. *Discourse Processes*, *38*, 1–30.

McNamara, D. S., Floyd, R. G., Best, R., & Louwerse, M. (2004). World knowledge driving young readers' comprehension difficulties. In Y. B. Yasmin, W. A., Sandoval, N. Enyedy, A. S. Nixon, & F. Herrera (Eds.), *Proceedings of the sixth international conference of the learning sciences: Embracing diversity in the learning sciences* (pp. 326–333). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–287.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43.

McNamara, D. S., & McDaniel, M. (2004). Suppressing irrelevant information: Knowledge activation or inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 465–482.

McNamara, D. S., & O'Reilly, T. (in press). Theories of comprehension skill: Knowledge and strategies versus capacity and suppression. In F. Columbus (Ed.), *Progress in experimental psychology research*. Hauppauge, NY: Nova Science.

Millis, K. K., Magliano, J. P., & Todaro, S. (2006). Measuring discourse-level processes with verbal protocols and latent semantic analysis. *Scientific Studies of Reading*, *10*, 251–283.

O'Reilly, T., & McNamara, D. S. (2006). Reversing the Reverse Cohesion Effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, *43*, 121–152.

Oakhill, J. (1994). Individual differences in children's text comprehension. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. New York: Academic.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford Press.

Perfetti, C. A. (1994). Psycholinguistics and reading ability. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 849–894). New York: Academic.

Shinn, M. R. (Ed.). (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford.

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.

Suh, S., & Trabasso, T. (1993). Inferences during reading: Converging evidence from discourse analysis, talk-aloud protocols, and recognition priming. *Journal of Memory and Language*, *32*, 279–301.

Trabasso, T., & Suh, S. (1993). Understanding text: Achieving explanatory coherence through online inferences and mental operations in working memory. *Discourse Processes*, *16*, 3–34.

Trabasso, T., van den Broek, P., & Suh, S. (1989). Logical necessity and transitivity of causal relations in the representation of stories. *Discourse Processes*, *12*, 1–25.

van den Bergh, H. (1990). On the construct validity of multiple-choice items for reading comprehension. *Applied Psychological Measurement*, *14*, 1–12.

van den Broek, R., Risden, K., & Husebye-Hartman, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353–374). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies in discourse comprehension*. New York: Academic.

Wiley, J., & Rayner, K. (2000). Effects of titles on the processing of text lexically ambiguous words: Evidence from eye movements. *Memory & Cognition*, *28*, 1011–1021.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*, 292–297.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. J*ournal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 386–397.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162–185.

APPENDIX

EXAMPLE READING COMPREHENSION TEXT AND QUESTIONS

Franco Dictatorship

The Franco dictatorship, lasting from 1936–1975, was one of the most oppressive periods in modern Spanish History. Franco took power in Spain after the Spanish Civil War in 1936. Supporters of the prior government, known as Republicans, included most workers, liberals, socialists, communists, and Basque and Catalan separatists. The Franco government labeled all political opposition as communists and used that to justify their harsh actions. In the first four years after the war, the government imprisoned hundreds of thousands of people and executed many thousands of others. The Franco government tracked people suspected of Republican sympathies and persecuted them for decades.

The dictatorship's main source of political support included the army, the Catholic Church, and the Falange, the Spanish National Movement. The common enemies were the socialist and communist movements in Spain. The army provided the dictatorship with security, while the Catholic Church and the National Movement gave Franco's rule a measure of legitimacy. As long as Franco openly opposed communism, the Church turned a blind eye to the dictatorship. To this day, many Spanish citizens who lived under the dictatorship have a distrust of the Catholic Church.

Franco, who sympathized with fascist ideas, was a great admirer of Adolf Hitler. Spanish industries were inefficient and the transportation system was largely in ruins, making mobilization for war difficult. Thus, Spain was unable to offer assistance to Germany. Spain was forced to adopt an official policy of neutrality during the war. Despite this, Spain sold valuable raw materials, such as steel, to some of the Axis powers. Spain emerged from the war politically and economically isolated. Many countries cut off diplomatic relations with Spain also.

Domestically, Franco's economic policies further isolated Spain and led to a disastrous period of economic stagnation. Franco believed that Spain could achieve economic recovery and growth through rigorous state regulation of the economy. Franco's government made few investments to rebuild the nation's shattered infrastructure, as well as his policies effectively deprived Spain of foreign investment. Agricultural output and industrial production languished, wages plummeted, and the black market flourished. High inflation and low wages defined the Spanish economic landscape. To make matters worse, Franco refused to seriously open the Spanish economy to foreign trade and investment.

Franco was forced to institute changes that ultimately weakened his government's grip on the country. The cabinet was reorganized in order to increase labor and business representation in the government. Industrial production boomed. Impoverished agricultural workers left the fields for better paying jobs in the city. Labor agitation increased, workers were dissatisfied and organized into unofficial trade unions to press for better pay, benefits, and working conditions. By the late 1960's and early 1970's, Spain was a society at odds with the aging Franco dictatorship. The dictatorship finally lost power in 1975.

1. When did Franco take power in Spain? TEXTBASE
2. Identify at least two enemies and supporters of Franco's government. TEXTBASE

3. Why would some people living in Spain today distrust the Catholic Church? SITUATION MODEL
4. Was Spain neutral during World War II? Why or why not? SITUATION MODEL
5. What did Spain sell to its allies during World War II? TEXTBASE
6. What did most countries do to Spain after World War II? TEXTBASE
7. What were the causes of the great period of economic stagnation that followed World War II? SITUATION MODEL
8. Why did Franco re-organize his Cabinet and what were the results of that reorganization? SITUATION MODEL
9. Near the end of Franco's rule, why did agricultural workers leave their fields and what were the consequences? SITUATION MODEL
10. When did the Franco Dictatorship loose power? TEXTBASE

6

Developing and Validating Instructionally Relevant Reading Competency Profiles Measured by the Critical Reading Section of the SAT Reasoning Test™

Arthur VanderVeen and Kristen Huff
The College Board

Mark Gierl
University of Alberta, Edmonton, Alberta, Candad

Danielle S. McNamara, Max Louwerse, and Art Graesser
University of Memphis

This study investigated whether the SAT Reasoning Test™ critical reading section would support categorizing students into reading competency profiles that could be matched to reading comprehension instructional treatments. We conducted task analyses of the reading comprehension processes required to successfully answer test items on the SAT® critical reading section and developed 5 reading comprehension skills categories. Multidimensional statistical analyses of examinee item response patterns confirmed that the SAT critical reading section

measures differential performance on 4 of the 5 skills categories: (a) Word Meaning, (b) Sentence Meaning, (c) Global Text Meaning, and (d) Pragmatic Meaning. Further research is proposed to investigate whether the skill categories can be combined into reading competency profiles that represent differential competencies in lower level, text-processing skills versus higher level, integrative skills.

As part of its commitment to help all students connect to college success, The College Board is conducting ongoing research to provide descriptive, instructionally relevant feedback to students taking the SAT Reasoning Test™.[1] The goal of this study was to identify the reading comprehension processes that are measured by the critical reading section of the SAT and thus lay the groundwork for future studies that will investigate whether valid and reliable information on these underlying component comprehension processes could be used to diagnose specific reading comprehension strengths and weaknesses. The ultimate goal of this research is to report meaningful and descriptive feedback for the SAT critical reading section that will enable school administrators, teachers, and parents to match students with appropriate instructional programs and thus prepare them for the sophisticated reading tasks they will encounter in college.

INSTRUCTIONALLY RELEVANT ASSESSMENTS

The College Board's initiative to provide instructionally valuable feedback from the SAT aligns with expanded notions of test validity that have emerged over the last decade as education stakeholders demand data from assessments that can be used to improve curriculum and instruction and respond to individual learning needs (Carnegie Council on Adolescent Development, 1989; Glaser, 1986; Griffin, 2001; Huff & Goodman, 2007; Linn, 1986; National Education Goals Panel, 1991; Nitko, 1989; Pellegrino, Baxter, & Glaser, 1999; Pellegrino, Jones, & Mitchell, 1999; Snow, 1989; Stiggins, 1994). Designing assessments to support instructional decisions has been a critical component of standards-based accountability models wherein criterion-based assessments measure student progress toward the intended learning outcomes articulated in academic content standards. Administrators and teachers attempting to help students advance toward proficiency on those criterion

[1]The College Board is a not-for-profit membership organization founded in 1900 by several U.S. colleges to simplify and make more equitable the college admissions process. The member colleges developed a common college entrance examination, which later became known as the SAT Reading Test.

measures need to know more than just how students are performing with respect to the overall population of students who have taken the test; they need instructionally valuable feedback on the specific content knowledge, skills, and strategies that should be targeted instructionally to help students improve. It is within this context that the College Board is conducting research on the viability of extracting more information than a total score (i.e., 200–800) from the SAT critical reading section to classify students into reading competency profiles that can be matched to appropriate remedial reading instructional programs.

The first step toward using the SAT critical reading section to categorize examinees into reading competency profiles was to identify the cognitive processes and knowledge structures required to respond to the test items. A cognitive model describing the processes and knowledge structures that underlie proficient performance both in the assessment context and in the nontest domain enables measurement specialists to reason inductively from observed performance on the assessment tasks to expected performance in the domain (Greeno, 1989; Mislevy, 1994, 1996; Mislevy & Riconscente, 2005; Nichols, 1994; Schum, 1987). Our reading competency profiles would provide this link between the underlying component comprehension processes measured by the test and critical features of authentic critical reading tasks. This translation from cognitive processes measured by test items to authentic domain performance would make it possible to align weakness in the underlying component processes with known reading comprehension strategy interventions.

The SAT was not originally designed to provide targeted, diagnostic feedback on specific underlying reading comprehension skills. As a result, the SAT exam specifications do not reflect a cognitive model for how the component reading comprehension processes measured by the test items contribute to overall performance on the critical reading section of the test. This is not unusual for many standardized assessments developed through conventional test development methods—tests like the SAT that were designed primarily to differentiate individuals in terms of their predicted performance on selected criterion variables without having articulated an explicit cognitive model that links item-response patterns to the knowledge structures and cognitive processes that underlie domain performance (Everson, 2004). Although an explicit cognitive model may not have been articulated in advance, inferred cognitive models can nevertheless be used to better articulate the meaning of test scores on existing tests. Post hoc assessment task analyses are used to articulate a test's underlying, implicit construct of domain performance. "If one starts with a well-grounded construct theory," wrote Messick (1989), "the whole enterprise is largely deductive in nature. … But often one starts with a dimly understood test or set of tasks and attempts to induce score meaning from an assorted array of the test's empirical relationships" (p. 49). The development of the PSAT/NMSQT Score Report *Plus*™ provides a precedent for this kind of analysis. Researchers conducted a post hoc task analysis of the PSAT/NMSQT to identify the underlying skills measured by the test. The Score Report *Plus* provides individualized feedback to students about specific skills on which they need to improve, based on students' item-response patterns (DiBello & Crone, 2001).

Inferring the implicit construct underlying student performance on the SAT critical reading section required that we identify likely strategies students use to answer test questions, define the text processing and reasoning processes entailed in those strategies, and hypothesize how differences in examinee competencies in those processes would produce differences in the overall critical reading score on the SAT. However, interpreting the single critical reading score on the SAT in terms of an explicit cognitive information-processing model presents a number of challenges. For instance, such an interpretation assumes that the assessment tasks on the test tap multiple dimensions of reading comprehension that can be differentiated, observed, measured, and interpreted. Long considered unidimensional, the construct of verbal reasoning measured by the critical reading (formerly verbal) section of the test is represented by a total score that ranks examinees along a single dimension of critical reading skill. The reliability of that total score is affected by the dimensionality of the assessment; assessing examinees on two or more dimensions interferes with ranking them along a single dimension. Tests designed to report a single, total score use psychometric procedures that suppress the effect of competing dimensions in order to support unidimensional ranking. Competing dimensions are considered measurement error and construct irrelevant. Items are written to reduce construct-irrelevant variance, and items that display evidence of construct-irrelevant variance are systematically eliminated from the test through detailed analysis of pretested items.

Researchers have questioned the degree to which the SAT critical reading section is unidimensional. An answer to this question is essential to determining whether psychologically meaningful and diagnostically useful information could be obtained from the SAT critical reading score. Despite the rather persistent assumption that the SAT critical read items measure a unidimensional construct, previous research on the dimensionality of the SAT is neither extensive nor conclusive on this topic. For instance, Cook, Dorans, and Eignor (1988) used confirmatory factor analysis to assess the dimensionality of the SAT-V section. The results of the factor analysis revealed that the SAT-V was slightly multidimensional. The authors concluded by calling for additional dimensionality studies on the SAT because this type of research "might yield diagnostics that could be used to arrive at more informed psychometric decisions about test specifications, and about the equating and scoring of the SAT (p. 40). Dorans and Lawrence (1999) also concluded that the SAT-V was multidimensional at both the item and test level of analysis (pp. 21–32). Most recently, Gierl, Tan, and Wang (2005) conducted a series of exploratory and confirmatory analyses using both parametric and non-parametric dimensionality procedures using items from the critical reading section of the SAT. They concluded that there is a multidimensional basis for test score inferences on the SAT, as both the exploratory and confirmatory dimensionality procedures provided consistent evidence to suggest that the critical reading items measure multiple dimensions. These three studies suggest that although the SAT critical reading section has been scaled as a unidimensional construct that can yield highly reliable total score rankings, the underlying construct of reading comprehension itself may indeed measure multiple dimensions. What remains to be thoroughly investigated is whether these component dimensions could be differentiated, measured, scaled, and scored to provide instructionally valuable information on multiple component processes and strategies underlying reading comprehension performance while simultaneously contributing to a single, reliable, total score representing verbal reasoning ability.

Multidimensional models assume that more than a single ability is necessary to account for examinee test performance on measures of reading comprehension ability. A multidimensional view of reading ability assumes that reading is a complex process, composed of many processes. As Hunt (1985) described:

> These range from the automatic, involuntary acts of lexical identification to the planned strategies people use to extract meaning from lengthy texts. There are individual differences in all of these processes. They combine to produce "verbal intelligence." (p. 55)

Hunt went on to argue, however, that

> There is no disagreement between the psychometricians' observation that verbal comprehension behaves, statistically, as if it were a unitary ability and the experimental psychologists' and linguists' contention that verbal comprehension can be broken down into its component processes. (1985, p. 55)

Early efforts to parse out the multiple cognitive dimensions underlying test performance on standardized tests of reading comprehension used a *cognitive correlates approach*, in which examinee scores on a standardized measure were compared to examinee performance on a battery of laboratory tasks designed to measure specific components of a cognitive information-processing model (Frederiksen, 1982; Pellegrino & Glaser, 1979). Respondents were typically divided into high- and low-ability groups as evaluated by the standardized measure. Between-group analyses were conducted that compared performance on tasks measuring component processes, such as stimulus encoding and matching, memory search, and response execution. Analyses were conducted to test whether differential performance on the standardized measure could be predicted by differential performance on measures of the component processes. A significant mean difference between high- and low-ability groups suggested that the component process was instrumental in test performance (Hunt, 1978; Hunt, Frost, & Lunnenborg, 1973; Hunt & Lansman, 1975; Hunt, Lunnenborg, & Lewis, 1975; Perfetti & Goldman, 1976; Perfetti & Hoagaboam, 1975).

Others have used a task analysis approach to identify directly the component processes that affect test performance (Anderson, 1982; Embretson & Wetzel, 1987; Gorin, 2005; Sheehan & Ginther, 2001; Sternberg, 1977; Woltz, 1988). Such an approach involves conducting detailed task analyses of the test items and developing a model of performance to explain individual differences in test performance. These analyses attempt to make explicit what may have been only vaguely articulated in the test- and item-level specifications, namely, the cognitive processes and knowledge structures required by the examinee to answer each question correctly. Detailed descriptions of the processes entailed in responding to assessment items are developed into an overall model of performance. Such a model includes linking performance on the specific component processes to differential models of student performance that capture the critical distinctions among basic, proficient, and advanced competency in the domain. If the total score on the test is a valid measure of domain competence, then examinee profiles comprising various combinations of the multiple underlying processes being measured by the test items can be linked to domain competence and be used for diagnostic test score inferences.

COGNITIVE TASK MODELING TO SUPPORT ITEM DIFFICULTY MODELING

We used a task analysis approach to develop reading competency profiles measured by the SAT. Task analysis studies have been conducted on other standardized tests, some of which are described later. Unfortunately, these studies have failed to provide instructionally valuable feedback, because most task analyses conducted to date have derived their model of reading comprehension from cognitive information-processing models that are only distally related to the classroom learning environment and the learning activities one would hope to inform (Pellegrino, Baxter, & Glaser, 1999). Instructionally relevant assessments should support the interpretation of test scores according to theories of cognition, learning, and instruction. The reading comprehension model used to guide the task analyses in this study was designed to align better with how students develop increasing competence in reading comprehension so that potential diagnostic feedback would better support decisions about instructional practice.

There are reasons why most task analysis studies to date have not reported student performance in ways that are easily translatable to classroom instructional practice. Most task analysis studies have attempted to identify those item features that best account for *variance in item difficulty*, defined as the probability of correct response. Items for which the probability of correct response increases sharply with small increases in estimated ability level effectively discriminate between examinees with similar, but not identical, ability levels; such items are critical for assessments designed to rank examinees. Many task analysis studies have, therefore, sought to identify interactions between student ability traits and item features that account for the greatest variance in item difficulty.

Embretson and Wetzel (1987), for example, applied a cognitive processing model of reading comprehension to account for variance in item difficulty on the Armed Services Vocational Aptitude Battery. The authors analyzed the test items and identified item features that would account for differences in the cognitive complexity of the processes required to answer the questions correctly. Embretson and Wetzel identified two general processes underlying performance on multiple-choice reading comprehension questions based on reading a passage: (a) text representation and (b) test item response decisions. Numerous studies involving verbal protocols have also shown that multiple-choice reading comprehension questions involve cognitive processes related both to constructing representations of text meaning and to decision processes related specifically to the cognitive processes involved in answering multiple-choice test items (Hannon & Daneman, 2001; Kirsch & Mosenthal, 1990; Mosenthal & Kirsch, 1991). Embretson and Wetzel found that under a variety of treatment conditions both processes account for item difficulty to varying degrees, depending on length of passage and format of test item.

The list of variables considered by Embretson and Wetzel (1987) illustrates the types of cognitive information-processing components typically considered when attempting to account for item difficulty: modifier propositional density, predicate propositional density, content word frequency, percentage content words, percentage relevant text, vocabulary level of the test item distracters, vocabulary level of the correct response, confirmation, falsification, reasoning required of the distracters, and reasoning required of the correct response. For instance, texts and questions with low-frequency words and propositionally dense texts increase cognitive demands and, therefore, item difficulty. *Percentage relevant text* refers to the amount of text in the passage that must be processed to answer the question; as the amount of required reading increases, so do demands on memory and encoding, increasing item difficulty. *Confirmation* and *falsification* refer to whether explicit information is available in the text to either confirm the correct answer or reject the distracter options; when such explicit information is not available, inference processes are required. *Reasoning* refers to the degree to which the syntactic structure and/or semantic content of the relevant text must be transformed to recognize its relevance to confirming or rejecting response options.[2]

As is evident, all of these item features are derived from a cognitive information-processing model specific to the assessment task involving multiple-choice reading comprehension questions. Embretson and Wetzel (1987) found that the majority of item difficulty variance on the Armed Services Vocational Aptitude Battery was accounted for by variables related to item decision processing, specifically, processes related to mapping information between passages and response alternatives. They also found that using low-frequency words in the correct response option made it less likely that examinees would take on the increased cognitive load required to process the correct option, thus increasing the difficulty of the test item. Although these item features may effectively predict item difficulty, our opinion is that curriculum and instructional designers would have difficulty diagnosing and classifying students into meaningful reading competency profiles based on these item decision processes because it is not clear how they relate to reading performance outside of the assessment context.

Although Embretson and Wetzel's (1987) model demonstrates the potential offered by cognitive approaches to test design, it is typical of most work to date that has focused on atomistic, discrete cognitive processes that can be isolated and manipulated to increase an item's discrimination power. The authors of the study presented in this chapter were conscious of how the selection of item attributes used to code the SAT items would constrain the types of inferences that could be supported by analyses of item response patterns. We thus selected attributes that better align with nontest reading performance so that test data would better support instructional decisions. Greeno (1976) captured this tension that researchers face when selecting the grain size of the attributes to include in their cognitive model:

[2]Reasoning is coded using a cognitive complexity scale developed by Anderson (1982).

> It may not be critical to distinguish between models differing in processing details if the details lack important implications for quality of student performance in instructional situations, or the ability of students to progress to further stages of knowledge and understanding. (p. 133)

The preceding discussion illustrates how the objective of ranking examinees on a single scale has constrained the way test developers model and measure reading comprehension (Snow & Lohman, 1989). The limited value of these efforts for diagnostic purposes suggests that a cognitively diagnostic model should describe examinees' solution strategies and cognitive processes in ways that will support generalizing to nontest domain performance and matching differential reading comprehension competencies to appropriate remedial reading instructional programs.

PSYCHOLINGUISTIC MODELING OF READING COMPREHENSION

During the 1980s, experimental psychologists in the reading research group at the Learning Research and Development Center at the University of Pittsburgh used a psycholinguistic approach to model reading comprehension in ways that more closely reflect actual reading comprehension activities outside the assessment context and account for individual differences in reading comprehension ability. On the basis of this research, Perfetti (1985a) developed his *verbal efficiency theory*, which represents reading ability as a set of interactive processes by which readers model the meaning of texts. Although this work is now 20 years old, we find it relevant to the purposes of this study, as Perfetti's model describes a hierarchy of multiple dimensions of reading comprehension that is anchored in theoretical models for how readers construct and integrate meaning from texts. Perfetti's dimensions provide a model, therefore, for defining item attributes that better align with actual reading behavior and may support instructionally relevant diagnostics.

According to the model, readers parse and assemble propositions into local representations of text meaning and then integrate these local representations into higher level representations. Local processes include recognizing words, assembling propositions, integrating propositions across sentences, and developing representative text models. Higher level, integrative processes include activating relevant schemas from prior knowledge to structure local propositions into meaningful text structures; making inferences to fill in gaps in the textbase and elaborate the situation model; integrating propositions across sentences and larger sections of text to construct macrostructures; and making inferences based on contextual discourse factors, such as author's purpose. Other higher level processes that affect reading ability include metacognition and comprehension monitoring which, together, enable readers to allocate processing resources strategically to increase comprehension (Perfetti, 1985b). Perfetti hypothesized that the lower level linguistic processing tasks were more resource intensive than the higher level integrative processes and, therefore, that ability with respect to the lower level linguistic processes would be the stronger predictor of general reading ability. Perfetti did not claim that higher level integrative processes were not significant in determining reading ability; instead, the point of his verbal efficiency theory was to highlight that local processes were so resource intensive that they represented the limiting factor and would be the best indicator of individual differences in reading ability. Perfetti and others found that differences in working memory capacity, lexical access, elementary proposition encoding, and proposition integration within and across sentences correlated significantly to differences in general reading ability (Daneman & Carpenter, 1980; Just & Carpenter, 1992; Kail, Chi, Ingram, & Danner, 1977; Lesgold & Perfetti, 1978; McClelland & Rumelhart, 1981; Perfetti & Goldman, 1976; Perfetti & Hogaboam, 1975).

On the basis of this work, Perfetti (1985a) identified four dimensions of reading comprehension processes that could be used to define profiles of general reading ability. These dimensions together make up the "text work" involved in constructing representations of text meaning: (a) lexical access (A), (b) proposition assembly (P), (c) proposition integration (I), and (d) local text modeling (M). *Lexical access* includes word recognition and recall. *Proposition assembly* refers to elementary encoding involving assembling a single proposition from only a few words. *Proposition integration* refers to higher level integrative and inferential processes required to integrate propositions within and across sentences or to fill in text gaps. *Text modeling* refers to developing a mental model of what is explicitly represented in the text. Explicit text models become the basis for the higher level processes involving making connections to existing knowledge structures and schemas that enable the interpretive, inferential, and critical comprehension of a text that goes beyond the text itself. The theory assumes that each of these distinct sets of processes, or components, produce differences in general reading ability.

Readers can be characterized by their efficiencies with respect to each of these components of text processing, and texts can be characterized by the cognitive load requirements for each component. For skilled readers and easy texts, lexical access (A), proposition assembly (P), and local text modeling

(M) are typically quite automatic; proposition integration (I) across sentences generally introduces the greatest cognitive demand. If total text effort = T, then profiles of texts and readers can be described by the relative magnitudes of the ratios of each component to total effort: A/T, P/T, I/T, and M/T. Text work components vary with the interaction of texts and individuals. The difficulty of a text can be described by the value of the cognitive processing load (measured by an appropriate time-response variable averaged across individuals) for each process; an individual's reading ability profile can then be described by his or her relative processing efficiencies in A, P, I, and M for that text (or corpus of texts). Text processing performance for a given text or corpus with determined text difficulty parameters for each process can be predicted for different reading ability profiles.

Perfetti's work (1985a) suggested that experimental researchers could define and validate profiles of reading ability based on component processes that are more closely related to actual reading comprehension activities. Assessments designed to identify item-response patterns that reflect the hypothesized ability profiles could be diagnostically useful for guiding instructional treatments. Multidimensional statistical analyses would enable investigators to relate the item-response patterns to the hypothesized ability profiles, differentiating individual performances in instructionally relevant ways. Instead of trying to derive interpretive information from an examinee's location on a unidimensional ability scale defined by item difficulty, such a multidimensional model would link item-response patterns to instructionally relevant reading competency profiles that reflect critical differences between successful and unsuccessful performance in the larger domain.

Perfetti and his colleagues (Perfetti, 1985a; Perfetti & Goldman, 1976; Perfetti & Hogaboam, 1975) have conducted numerous studies relating differences in general reading ability to various text processing components. They have focused especially on the contributions of lower level text processing skills, such as decoding orthographic structures, lexical access, semantic coding, and proposition assembly and integration. Multiple studies have shown that differences in lexical access and proposition assembly and integration correlate to differences in overall reading ability, supporting the inference that inefficiencies in lower level processes diminish overall reading performance (Perfetti, 1985b). Perfetti did not hypothesize or validate the exact relationship between these variables, however, and thus stopped short of articulating a comprehensive model for predicting how readers with differing competencies on the component processes will perform in observably different ways in nontest reading comprehension tasks.

TASK ANALYSES OF THE SAT CRITICAL READING SECTION

Our task analysis of the SAT critical reading section is consistent with the multidimensional analysis advocated by Magliano et al. (chap. 5, this volume). Our goal was to develop profiles of reader proficiencies that would provide a basis for matching the strengths and weaknesses of readers as measured by the SAT critical reading section to appropriate reading comprehension instructional treatments. We analyzed the critical reading test items to determine the products and processes of comprehension required to answer the items correctly.

Any attempt to make the SAT critical reading section more cognitively diagnostic must support interpretations of student test performance in terms of instructionally relevant reading competency profiles while also reliably ranking examinees on a single scale of reading proficiency. This constraining requirement that the SAT critical reading section continue to reliably differentiate examinees on a single scale limits the degree to which test items can be modified to support diagnostic feedback. Research efforts to support these related but different objectives began at the College Board and Educational Testing Service in the 1990s as researchers analyzed the assessment tasks on the PSAT/NMSQT (DiBello, 2002; DiBello & Crone, 2001). The research team developed and elaborated the rule-space method to identify and report on component dimensions of reading comprehension performance as measured by the test (K. K. Tatsuoka, 1983, 1984, 1985, 1990, 1991; K. K. Tatsuoka & Tatsuoka, 1991; M. M. Tatsuoka & Tatsuoka, 1987; DiBello, Stout, & Roussos, 1995; Buck et al., 1998). The PSAT/NMSQT score report now includes suggestions for how each examinee could improve on up to three dimensions of reading comprehension as measured by the critical reading section of the test. Test specifications have not been modified to support this reporting on distinct skills measured by the test; instead, the measurement and reporting mechanisms used to identify reading skills needing improvement are based on a task analysis of the existing test.

The College Board began a similar effort in 2003 to conduct a task analysis of the critical reading section of the new SAT. The central question was whether examinee performance on clusters of items coded as measuring component processes of reading comprehension could support differentiating examinees into instructionally relevant reading competency profiles (Huff, 2004; VanderVeen, 2004). Could student performance on clustered items support diagnostic inferences while the total score still maintained the test's reliability and predictive validity? Answering these questions required articulating a model that would account for differences in general reading proficiency according to measured differences in component processes. Moreover, the component processes contributing to variation in general reading proficiency must be defined in terms that could be matched to reading comprehension interventions. A team of reading comprehension researchers specified such a model and iterated multiple times through an item-coding process on 11 different test forms of the SAT to see whether the existing test items measured the component processes. The researchers refined the model through each iteration, converging on a model that was both consistent with the theoretical construct and supported by the existing test items. The researchers then coded the items in the critical reading section of the prototype form of the new SAT, which was administered to more than 45,000 students in the nSAT and new Preliminary SAT/National Merit Scholarship Qualifying Test (nPSAT/NMSQT) Spring 2003 field trial (Liu, Feigenbaum, & Walker, 2004). Data from the field trial were then analyzed using multidimensional statistical analyses to see whether item response patterns generated by the test could be interpreted to fit with the hypothesized reading ability profiles (Gierl, Tan, & Wang, 2005; Gierl, Leighton, et al., 2005).

Unlike task analysis studies designed to account for variations in item difficulty, the coding schema used for this study did not code on comprehension processes related to item processing. We felt that processes related to item processing did not generalize well to explain performance outside the assessment context, defined as the kinds of reading comprehension tasks students would face in college. Instead, we sought to identify both lower level linguistic processes and higher level integrative and interpretive processes that would both differentiate high- and low-ability readers and support the classification of readers into instructionally relevant profiles that could be linked to appropriate remedial reading interventions.

The item attributes identified in this study represent, therefore, larger grained dimensions that could more easily align with appropriate instructional interventions. These larger grained dimensions can also be measured reliably by the limited number (67) of test items on the SAT critical reading section. Test items were coded on multiple dimensions, but only a single dominant dimension was used in our statistical analyses. A team of four expert raters coded the items independently, and a dominant dimension was identified through consensus. Interrater reliability for individual ratings was very high (.91; Gierl, Tan, & Wang, 2005).[3] Through multiple and iterative coding rounds, we identified five text processing skills measured by the

[3]We conducted a generalizability study (G-study) to determine the interrater reliability of the codings; 69.55% of the total variance was associated with the items. The variance component for raters accounted for only 1.59% of the total variance, indicating high rater consistency. The variance component associated with the residual was also high (28.86%); this contained the Rater × Item interaction confounded with other systematic sources of variance and random error.

critical reading section of the new SAT: (a) determining the meaning of words (Word Meaning); (b) understanding the content, form, and function of sentences (Sentence Meaning); (c) understanding the situation implied by a text (Situation Model); (d) understanding the content, form, and function of larger sections of text (Global Text Meaning); and (e) analyzing authors' purposes, goals, and strategies (Pragmatic Meaning; see Table 6.1 for detailed descriptions of each category). After additional coding, we reduced our five dimensions to four, because there were insufficient items to support reliable measurement on the skill category for Situation Model. Our subsequent analyses included only the four remaining skill categories.

Several differences between these categories and the component processes in Perfetti's (1985a) model are immediately apparent: Whereas Perfetti sought to account for differences in individual reading ability by measuring differences in the efficiency of lower level linguistic processes, the SAT has historically focused on higher level integrative, inferential, interpretive, and critical comprehension abilities, because these have been shown to be most predictive for college success. As such, SAT reading comprehension test items emphasize measurement of these skills, typically in combination. One item type, for example, asks examinees to infer two authors' points of view on a common issue represented in two separate passages, each approximately 600 words in length. Examinees must then reason through how one author would likely respond to a specific claim made by the other author. Such an assessment task clearly draws on multiple higher level integrative and inferential processes, as the examinee must summarize each author's position and then reason through a highly inferential process based on an implied situation model involving the common issue—each author's point of view—and other discourse-level factors implied by things such as tone, credibility, and author's purpose.

Although the goal of a task analysis is to identify and differentiate the critical component processes tapped by each assessment task, the goal of articulating reading ability profiles as measured by the SAT also calls for combining the discrete component processes into sets of comprehension activities that reflect text comprehension processes in the nontest domain. Achieving the strong emphasis on reasoning required by the test specifications frequently depends on tasks that involve an interactive interdependency among processing levels: Measures of lexical access, for example, typically require the analysis of contextual clues to differentiate between possible primary and secondary word meanings. Reasoning processes are tapped as examinees must check possible lexical meanings of words against the semantic case analyses involved in proposition assembly and the emerging meaning representation resulting from propositional integration. All of this is

TABLE 6.1
Skill Categories

Skill Category	*Description*	*Comment*
Determining the Meaning of Words (Word Meaning)	Student determines the meaning of words in context by recognizing known words and connecting them to prior vocabulary knowledge. Student uses a variety of skills to determine the meaning of unfamiliar words, including pronouncing words to trigger recognition, searching for related words with similar meanings, and analyzing prefixes, roots, and suffixes.	This skill category includes more than just lexical access, as word identification and lexical recall are combined with morphological analyses.
Understanding the Content, Form, and Function of Sentences (Sentence Meaning)	Student builds upon an understanding of words and phrases to determine the meaning of a sentence. Student analyzes sentence structures and draws on an understanding of grammar rules to determine how the parts of speech in a sentence operate together to support the overall meaning. Student confirms that his or her understanding of a sentence makes sense in relationship to previous sentences, personal experience, and general knowledge of the world.	This skill category focuses on the syntactical, grammatical, and semantic case analyses that support elementary proposition encoding and integration of propositions across contiguous sentences.
Understanding the Situation Implied by a Text (Situation Model)	Student develops a mental model (i.e., image, conception) of the people, things, setting, actions, ideas, and events in a text. Student draws on personal experience and world knowledge to infer cause-and-effect relationships between actions and events to fill in additional information needed to understand the situation implied by the text.	This skill category is a hybrid of the explicit text model and the elaborated situation model described by Kintsch (1998). As such, category three combines both lower-level explicit text interpretation and higher-level inferential processes that connect the explicit text to existing knowledge structures and schemata.

(*continued*)

TABLE 6.1 (*continued*)

Skill Category	*Description*	*Comment*
Understanding the Content, Form, and Function of Larger Sections of Text (Global Text Meaning)	Student synthesizes the meaning of multiple sentences into an understanding of paragraphs or larger sections of texts. Student recognizes a text's organizational structure and uses that organization to guide his or her reading. Student can identify the main point of, summarize, characterize, or evaluate the meaning of larger sections of text. Student can identify underlying assumptions in a text, recognize implied consequences, and draw conclusions from a text.	This skill category focuses on the integration of local propositions into macro-level text structures (Kintsch & van Dijk, 1978) and more global themes (Louwerse & Van Peer, 2003). It also includes elaborative inferencing that supports interpretation and critical comprehension, such as identifying assumptions, causes, and consequence and drawing conclusions at the level of the situation model.
Analyzing Authors' Purposes, Goals, and Strategies (Pragmatic Meaning)	Student identifies an author's intended audience and purposes for writing. Student analyzes an author's choices regarding content, organization, style, and genre, evaluating how those choices support the author's purpose and are appropriate for the intended audience and situation.	This skill category includes contextual and pragmatic discourse analyses that support interpretation of texts in light of inferred authorial intentions and strategies.

combined into a single item, making it difficult to code items according to discrete, atomistic cognitive processes. The skill categories identified through our task analyses reflect this tendency of the items to combine component processes from multiple levels.

EVALUATING MODEL FIT

We conducted extensive parametric and nonparametric dimensionality analyses using data from the critical reading section of the 2003 field trial to evaluate whether the coded dimensions provided a good fit to item response patterns. Data from Design 1 of the nSAT and nPSAT/NMSQT Spring 2003 field trial (Liu et al., 2004) were analyzed. More than 45,000 students from 679 high schools (primarily juniors) participated in the field trial. These students were from both public and private schools across rural, suburban, and urban areas and represented every geographic region in the United States. The design included 13 booklets containing different combinations of test sections taken from the then-current SAT, the new SAT (first administered in March 2005), and the new PSAT/NMSQT (first administered in October 2004). The booklets were spiraled within classrooms to achieve comparable groups and appropriate sample sizes for all follow-up analyses. Data from three different booklets were used in this study (Books 2a, 2c, and 5) in order to cross-validate our findings based on analyses of the primary sample. The sample was deemed to be similar to but not entirely representative of the baseline cohort of college-bound seniors who took the SAT in 2002. Nevertheless, the results from comprehensive analyses allowed Liu et al. (2004) to conclude that the sample would allow researchers to adequately evaluate important psychometric issues on the new SAT, including the dimensionality of the test.

We used multidimensional statistical analyses of the item data to determine whether the coded skill categories offered a valid model for describing dimensional differences in reading competency as reflected in the item-response patterns (Gierl, Tan, & Wang, 2005; Gierl, Leighton, et al., 2005). Confirmatory DIMTEST (Froelich, 2000; Froelich & Habing, 2001) analyses were used to test whether item-response patterns emerged that supported the hypothesized dimensions reflected in the four skill categories. Two additional variables were selected from the test specifications and included in the analyses, namely, item type and passage. Confirmatory DIMTEST analyses code and partition the test items into clusters based on hypothesized dimensions and measure the covariance between scores on each cluster for examinees with the same score on the primary cluster. After conditioning on scores associated with the cluster measuring the primary dimension θ_1 (i.e., the partitioning subtest, PT), the expected value of the covariance for those items hypothesized to measure a secondary dimension, θ_2 (i.e., the assessment subtest, AT) will be zero if AT and PT are measuring the same dimension. This outcome occurs because any dependency between AT item responses is removed by conditioning on the PT, if the test structure is truly unidimensional. The analyses based on the four skill categories, on item type, and on passage were compared to determine whether item-response patterns could best be described by some combination of skill category and item type or passage type.

RESULTS FOR TEST SPECIFICATIONS

The critical reading section contains 67 test items administered in two 25-minute sections and one 20-minute section. According to the test specifications, the primary dimension measured by the SAT critical reading section is verbal reasoning, which includes comprehension of text (words, sentences, and larger passages); drawing inferences; drawing implications and conclusions; reasoning through an argument, analogy, or situation model; identifying author's primary purpose; synthesizing information; and analyzing rhetorical strategies. The test is designed to measure this primary dimension across two item types (sentence completion items and passage-based reading items) and four content categories (Humanities, Social Studies, Natural Sciences, and Literary Fiction). This design requires examinees to apply their verbal reasoning skills in meaningful ways within the context of understanding sentences and reading passages drawn from the specified subject areas, rather than abstractly.

Using item format as the organizing principle, confirmatory DIMTEST showed that sentence completion and passage-based item types measure distinct dimensions of reading comprehension proficiency (see Table 6.2). Sentence completion items are intended to measure examinee ability to reason through the syntactic and semantic case structures of a sentence to construct a coherent local textbase. This includes vocabulary knowledge. Passage-based reading items are designed to measure higher order reading skills, including constructing meaning representations of larger sections of text (e.g., main idea, primary purpose, relationships between paired passages); drawing inferences and implications, frequently about an author's views; inferring an author's rhetorical strategies and purposes as reflected in specific textual choices; and using analogical reasoning to apply an idea or relationship illustrated in the text to a different context. The confirmatory DIMTEST results suggest that these different item types are indeed measuring distinct dimensions of a larger verbal reasoning construct.

The dimensions measured by the passage-based item types are further differentiated by specific passage. Using reading passage as the organizing principle, confirmatory DIMTEST showed that items associated with each reading passage measure a distinct dimension (see Table 6.3). This result is consistent with previous research indicating that reading comprehension performance is influenced by passage topic (Gierl, Tan, & Wang, 2005; Gierl, Leighton, et al., 2005) and that passages tend to assess distinct dimensions (Bolt, 2005; Gierl, 2004; Stout et al., 1996).

TABLE 6.2

Confirmatory DIMTEST Results by Item Type

Book	T	P
2a	6.0065	0.0000
2c	9.3176	0.0000
5	7.8681	0.0000

Note. The sentence completion items served as assessment subtest (AT) and the critical reading items served as the partitioning subtest (PT).

TABLE 6.3

Confirmatory DIMTEST Results by Reading Passage

	Passage 1		*Passage 2*		*Passage 3*		*Passage 4*	
Book	*T*	*P*	*T*	*P*	*T*	*P*	*T*	*p*
2a	2.9267	0.0017	2.9879	0.0014	3.3044	0.0005	6.0956	0.0000
2c	3.2240	0.0006	1.8548	0.0318	4.0946	0.0000	3.5783	0.0002
5	2.7549	0.0029	1.8719	0.0306	6.1057	0.0000	5.0210	0.0000
	Passage 5		*Passage 6*		*Passage 7*		*Passage 8*	
Book	*T*	*P*	*T*	*P*	*T*	*P*	*T*	*p*
2a	0.2768	0.6090	2.4242	0.0077	4.2212	0.0000	3.8733	0.0001
2c	0.2164	0.4143	3.1612	0.0008	3.6696	0.0001	7.5926	0.0000
5	0.7026	0.2411	1.7746	0.0380	4.0021	0.0000	6.2434	0.0000

Note. Each reading passage served as a separate AT with the remaining items serving as PT.

RESULTS FOR SKILL CATEGORIES

The second set of confirmatory analyses used the cognitive skill categories identified through this study as the organizing principle to guide the analysis of dimensionality (see Table 6.4). Confirmatory analyses showed that Word Meaning and Sentence Meaning each measure distinct dimensions of reading comprehension proficiency for all samples and that Pragmatic Meaning measures a distinct dimension for one sample.

As shown earlier, item type produced dimensionally distinct clusters when the test specifications were used as the organizing principle, so cognitive skills and item format were crossed and the dimensionality of the resulting clusters was evaluated (see Table 6.5). Confirmatory analyses showed that for all three

TABLE 6.4
Confirmatory DIMTEST Result by Skill Category

	Word Meaning		*Sentence Meaning*		*Global Text Meaning*		*Pragmatic Meaning*	
Book	*T*	*p*	*T*	*p*	*T*	*p*	*T*	*p*
2a	4.7681	0.0000	0.2039	0.4192	2.5645	0.0052	0.3646	0.3577
2c	5.1321	0.0000	0.3301	0.6293	3.7985	0.0001	2.1892	0.0143
5	4.7999	0.0000	1.1444	0.1262	2.6066	0.0046	0.3344	0.6310

Note. Each dimension served as a separate AT with the remaining items serving as PT.

samples sentence completion items coded to Word Meaning and sentence completion items coded to Sentence Meaning each measured distinct dimensions of reading comprehension proficiency. (Items were coded to a single dominant skill category, so these item clusters did not overlap.) Confirmatory analyses showed that for all three samples passage-based reading items coded to Sentence Meaning and passage-based reading items coded to Global Text Meaning each measured distinct dimensions of reading comprehension proficiency. Analyses also found that for one sample passage-based reading items coded to Pragmatic Meaning measured a distinct dimension of reading comprehension proficiency. Thus, when skills categories are crossed with item types as described earlier, Word Meaning and Sentence Meaning are each dimensionally distinct for sentence completion items, and Sentence Meaning and Global Text Meaning are each dimensionally distinct for passage-based reading items.

MULTIDIMENSIONALITY-BASED DIF ANALYSES

We next set out to confirm these results by conducting multidimensionality based differential item functioning (DIF) analyses. DIF is an important analytical tool for relating primary and secondary dimensions measured by tests (cf. Gierl, Bisanz, Bisanz, & Boughton, 2003). DIF occurs in an item when examinees of equal abilities on the primary trait (or traits) measured by a test but from different populations (e.g., males vs. females, native English speakers vs. English language learners, different ethnic subgroups) differ in their probability of answering the item correctly. Items that display DIF are believed to measure at least one secondary dimension that offers a sizable advantage or disadvantage to a particular subgroup that is otherwise equal in ability on the primary trait measured by the test. An example would be an item intended to measure logical reasoning skills that asks examinees to reason through a logical problem posed within the context of the rules of a football game. Such an item would likely favor males over females because, on average, males are

TABLE 6.5
Confirmatory DIMTEST Result by Item Type and Skill Category

	Sentence Completion						*Passage-Based Reading*					
Skill	*2a*		*2c*		*5*		*2a*		*2c*		*5*	
	T	*p*	T	p	T	P	T	p	T	P	T	p
Word Meaning	5.2510	0.0000	5.8057	0.0000	4.8734	0.0000	–0.0427	0.5170	–0.1366	0.5543	0.5912	0.2772
Sentence Meaning	3.3576	0.0004	2.4194	0.0078	4.0147	0.0000	1.9613	0.0249	2.4984	0.0062	2.7293	0.0032
Global Text Meaning	NA	NA	NA	NA	NA	NA	2.4921	0.0063	3.8248	0.0001	2.5599	0.0052
Pragmatic Meaning	NA	NA	NA	NA	NA	NA	0.4368	0.3311	2.2251	0.0130	-0.2865	0.6128

Note. Each dimension served as a separate AT with the remaining items serving as PT. NA means that the cell contained 0 items.

more familiar with such rules, offering them a richer set of context-specific knowledge structures that would facilitate their reasoning processes. Analyzing the characteristics of test items that produce DIF across subgroups can, therefore, yield information about the dimensionality of the test.

As noted earlier, the critical reading section is designed to have examinees demonstrate their verbal reasoning skills within the context of understanding sentences and reading passages drawn from specified subject areas, rather than abstractly. Although measuring the primary dimension within meaningful reading comprehension contexts increases the validity of the measurement and the score interpretations, such contextual factors introduce secondary dimensions being measured by the test that can create problems with respect to DIF. Students who have taken more courses in one or more of the targeted content areas will likely have an advantage on items based on passages drawn from that area. If subgroups of the test-taking population systematically take more courses in one of the content areas (e.g., if males, on average, take more science courses in high school), then items based on passages drawn from that subject area will likely display DIF characteristics favoring that subgroup. Indeed, O'Neill and McPeek (1993) reported that reading comprehension items tend to favor males when the content has to do with science and to favor females when the content has to do with the humanities. Test developers examine these DIF factors and evaluate whether the secondary dimensions producing the DIF characteristics (e.g., the ability to understand science passages) are central to or tangential to the primary construct (verbal reasoning in context) and whether the resulting DIF is an unfair advantage or simply the result of summary characteristics of the subgroup (course-taking behaviors) that produce differential performance on the overall construct. Efforts are taken to eliminate DIF-producing items that measure secondary dimensions that are irrelevant to the primary construct.

A multidimensional model for DIF links substantive item characteristics, such as the reading comprehension processes identified through the task analysis in this study, to differential item response patterns observed across subgroups. If bundles of items coded to a skill category produce differential performance patterns for examinees of equal ability across subgroups, then we can infer that the items indeed include distinct secondary dimensions that are interacting differentially with the different subgroups. We may hypothesize why the secondary dimensions interact differentially with the different subgroups and use confirmatory statistical analyses to determine whether our proposed multidimensional construct provides a good fit with the data. A good-fit model will further validate our model.

We conducted differential item and bundle functioning (DIF and DBF) analyses to evaluate whether the skill categories based on the component processes underlying the skill categories identified in this study produced differential performance for examinees of equal ability across subgroups. Our analyses were based on the Shealy–Stout multidimensional model for DIF (Shealy & Stout, 1993), which postulates that DIF is caused by the presence of multidimensionality. SIBTEST was used to conduct the DIF and DBF analyses. If the test items measuring a specific skill category functioned differentially against one group of examinees in a consistent manner, then it would be reasonable to conclude that this cognitive skill constitutes a secondary dimension the SAT is measuring. In other words, the focus in our DIF analysis was to identify auxiliary secondary dimensions that elicit group differences.

We conducted three DIF/DBF studies: (a) gender (males vs. females), (b) language (English as a first language vs. English as a second language), and (c) ethnicity (White vs. non-White). For the gender DIF/DBF analyses, males served as the reference group; for the language DIF/DBF analysis, native English-speaking examinees served as the reference group; for the ethnicity DIF/DBF analysis, White examinees served as the reference group. For all statistical analyses, positive DIF/DBF values favor the reference group and negative DIF/DBF values favor the focal group. In each analysis, the studied subtest includes the items associated with one primary skill category, whereas the matching subtest included the remaining test items.

Although analyses were conducted at both the item level and for bundles of items coded to the skill categories, we present only the results for the item bundles here. The DBF analysis is designed to test whether the item bundles, organized according to skill category, elicit group differences. The results for the three DBF studies are presented in Table 6.6. These outcomes are even more apparent when the items are plotted by primary skill category (see Figures 6.1–6.3).

For the gender DBF analysis, three of the four bundles displayed DBF; that is, Word Meaning, Global Text Meaning, and Pragmatic Meaning displayed DBF ($p < .05$), with Word Meaning favoring males and Global Text Meaning and Pragmatic Meaning favoring females. For the language DBF analysis, Sentence Meaning systematically favored native English-speaking examinees. For the ethnicity DBF analysis, Sentence Meaning and Global Text Meaning displayed DBF, with Sentence Meaning favoring White examinees and Global Text Meaning favoring non-White examinees.

Taken together, results from the DIF and DBF analyses reveal that the skill categories tap auxiliary secondary dimensions that can elicit some group differences. All four primary skills systematically favored one group in at least one of the three subgroup analyses. These confirmatory multidimensionality based DIF and DBF analyses suggest that latent variables specific to each group are made apparent when group performance on the item bundles coded

TABLE 6.6
Differential Bundle Functioning Results from SIBTEST for the Three DIF Studies in Critical Reading

		Word Meaning		*Sentence Meaning*		*Global Text Meaning*		*Pragmatic Meaning*	
	Book	β_{UNI}	p	β_{UNI}	p	β_{UNI}	p	β_{UNI}	p
Gender DIF	2a	0.602	0.000	0.115	0.298	-0.625	0.000	-0.234	0.000
	2c	0.633	0.000	-0.040	0.724	-0.596	0.000	-0.132	0.036
	5	0.613	0.000	0.124	0.292	-0.700	0.000	-0.163	0.009
Language DIF	2a	-0.089	0.490	0.846	0.000	-0.154	0.330	0.117	0.232
	2c	-0.241	0.107	0.808	0.000	-0.195	0.289	0.052	0.598
	5	0.047	0.732	0.555	0.004	0.140	0.450	-0.138	0.165
Ethnicity DIF	2a	-0.132	0.138	0.568	0.000	-0.360	0.001	0.125	0.057
	2c	0.002	0.986	0.631	0.000	-0.237	0.057	0.017	0.817
	5	-0.084	0.389	0.624	0.000	-0.156	0.244	0.053	0.461

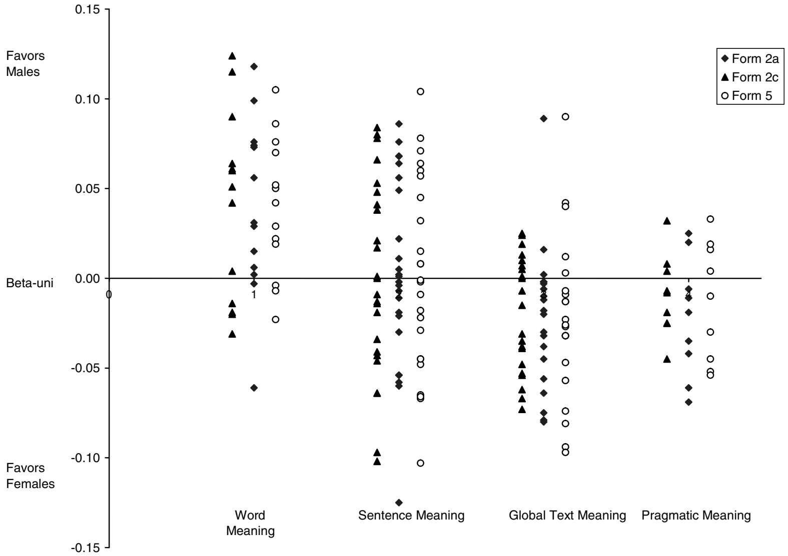


Figure 6.1. Differential bundle functioning graphical results using gender as the grouping variable in critical reading.

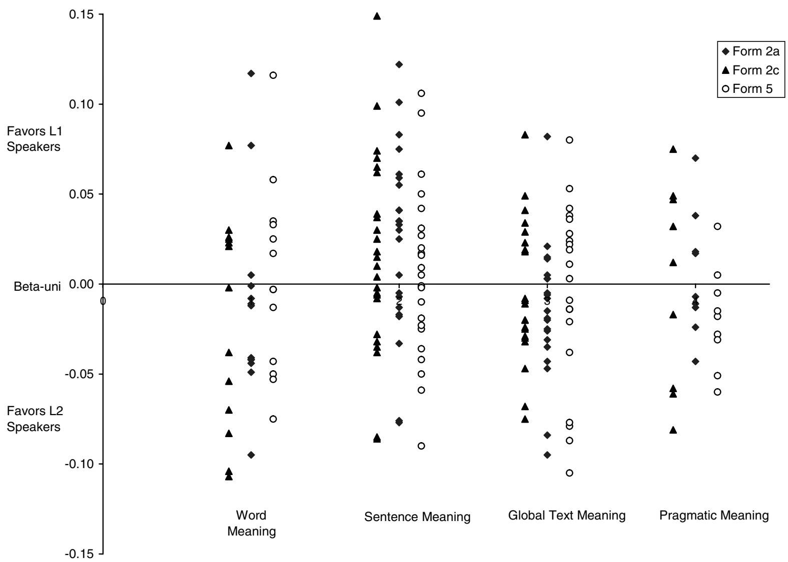


Figure 6.2. Differential bundle functioning graphical results using language as the grouping variable in critical reading. L1 = English as a first language; L2 = English as a second language.

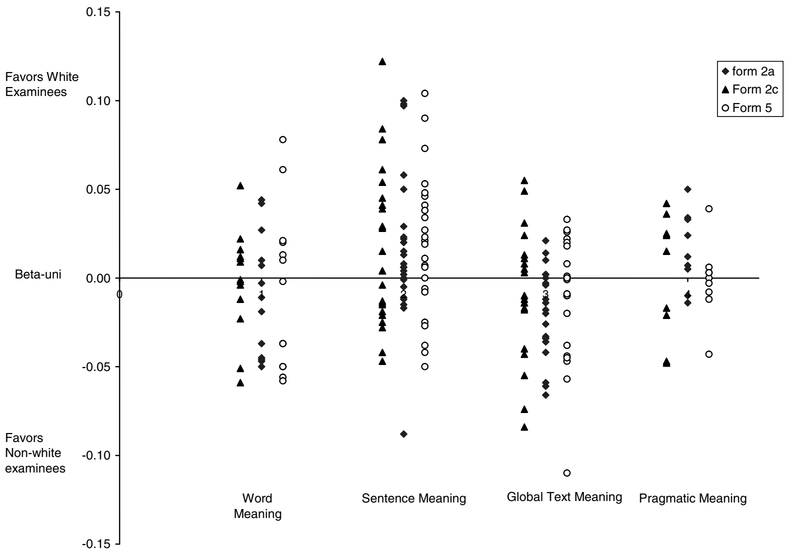


Figure 6.3. Differential bundle functioning graphical results using ethnicity as the grouping variable in critical reading.

to each skill category is measured. Differential group performance within a skill category suggests that the focal and reference groups differ on a latent trait (or traits) that affect performance on this dimension. Although our analyses do not hypothesize and validate what these latent variables may be for each group, their presence can be inferred through the DIF/DBF analyses. The fact that the skill categories define a coherent dimension apparently affected by latent explanatory variables provides further statistical validation for our model.

If we conclude that these secondary dimensions producing the DIF and DBF characteristics constitute component processes that are central to the primary construct (verbal reasoning in context), and not nuisance dimensions that unfairly advantage or disadvantage specific subgroups, we have further validation of the skill categories and will have confidence using them to define a set of reading ability profiles that can guide instruction. Although DIF should be minimized when it represents the unfair differential functioning of a secondary dimension that is not relevant to the construct, its presence is not necessarily an indicator of an irrelevant or unfair interaction between a secondary dimension and a favored subgroup. Because the skill categories are based on the reading processes identified through the task analyses, and not on the subject-specific knowledge structures associated with the passage content areas, we have confidence that the DIF- and DBF-producing characteristics of the test items are measuring construct-relevant auxiliary dimensions that ought to be included in the assessment.

DEFINING READING COMPETENCY PROFILES BASED ON COMPONENT DIMENSIONS

Identifying component processes of reading comprehension measured by the SAT is a first step toward providing feedback on student performance in ways that could be instructionally valuable. We infer that these components of performance, apparent in the item response data in the dimensionality analyses, are due to individual differences among the examinees in the reading comprehension processes identified through the task analyses and used to code the test items. To inform educational practice, cognitive diagnostic assessments must also relate these component processes to observable differences in examinee nontest reading comprehension performance. Predicting reading performance based on individual differences in component comprehension skills requires articulation of an explicit model for how different configurations of these component skills will likely produce different observable reading behaviors. We present the reading profiles that follow as initial attempts to define student models based on the component reading skills identified through this study. Subsequent research is planned to validate whether these student models predict differential reading performance, including differences in the total critical reading score and alternative measures of overall reading proficiency.

Much work has been done to measure differences in lower level linguistic processing, but less is known about how differential competencies in higher level reading processes (e.g., integration of propositions, inferencing, schema activation, and text modeling) contribute to differences in general reading performance. Because the SAT critical reading section focuses primarily on these higher level reading processes, further research into how these processes contribute differentially to reading performance is necessary if the SAT is to provide diagnostic feedback to examinees and educators. This study identified four component reading comprehension skills measured by the SAT, which span from lower level processes associated with lexical and semantic access, syntactic parsing, and proposition assembly to higher level integration and text modeling processes.

Using combinations of student skills or attributes that span local and global text processes is challenging, because reading comprehension processes are highly interdependent. Assessment tasks intended to measure higher level integrative or inferential processes also depend on lower level linguistic processes. Readers with deficits in sentence-level processing may have access to a rich network of schema-based knowledge structures or metacognitive comprehension monitoring skills that enable them to compensate for weak skills in lower level linguistic processes, such as syntactic analysis. We make some hypothesized predications on how different configurations of student attributes with respect to local and global text processing may produce consistent patterns in observable reading performance behaviors. These predictions will need to be refined through additional research, especially with respect to differentiating the relative contributions of local and global text processing skills to overall reading performance.

The profiles that follow are similar in intent as those defined by Perfetti (1985a), except that the skill categories developed through this study include both the lower level linguistic processes analyzed by Perfetti and the higher level integrative, inferential, and interpretive processes tapped by the complex reading comprehension tasks on the SAT. With regard to the lower level linguistic processes, we assume that high-quality lexical and semantic access (Word Meaning) facilitates all other linguistic and text-modeling processes. We also assume that syntactic processing and semantic case assignment (Sentence Meaning) contribute to differential reading ability through their facilitation of proposition assembly and integration processes. With regard to the higher level text-modeling processes, we assume that access to schema-based knowledge structures—including knowledge of story grammars for narrative texts and knowledge of typical organizational patterns for expository texts (e.g., cause and effect, problem solution, claim and evidence)—facilitates text modeling processes, including connecting the text model to an implied situation model, integrating local propositions into summary macrostructures, and drawing elaborative inferences and critical interpretations that go beyond the text (Global Text Meaning and Pragmatic Meaning). Given these assumptions, we propose the following reading competency profiles based on the four skill categories identified through this study:

Profile 1: Examinees have strong lower level linguistic skills (lexical and semantic access and syntax analysis skills) and strong higher level text-modeling skills (access to knowledge-based schemas and text structures and strong skills for analyzing pragmatic variables, e.g., author's purpose, intended audience, and the communication context).

Profile 2: Examinees have strong lower level linguistic skills (lexical and semantic access and syntax analysis skills) but weak higher level text-modeling skills (limited access to knowledge-based schemas and text structures and weak skills for analyzing pragmatic variables, e.g., author's purpose, intended audience, and the communication context).

TABLE 6.7
Text Processing Levels by Reading Comprehension Profiles

		Lower Level Processes	
		Strong	*Weak*
Higher Level Processes	*Strong*	Profile 1	Profile 3
	Weak	Profile 2	Profile 4

Profile 3: Examinees have weak lower level linguistic skills (lexical and semantic access and syntax analysis skills) but strong higher level text-modeling skills (access to knowledge-based schemas and text structures and strong skills for analyzing pragmatic variables, e.g., author's purpose, intended audience, and the communication context).

Profile 4: Examinees have weak lower level linguistic skills (lexical and semantic access and syntax analysis skills) and weak higher level text-modeling skills (limited access to knowledge-based schemas and text structures and weak skills for analyzing pragmatic variables, e.g., author's purpose, intended audience, and the communication context).

Table 6.7 illustrates how the profiles combine lower level and higher level processes.

DISCUSSION AND CONCLUSIONS

Given that large-scale, norm-referenced tests were not designed for the purpose of cognitively diagnostic assessment, it is a significant achievement that there is empirical evidence to support the interpretations of secondary dimensions on the SAT critical reading test. It is also a significant achievement that these dimensions reflect a psycholinguistic model of reading comprehension, as described by the four hypothesized skill profiles in the previous section. Next steps for research include conducting additional studies to further refine the hypothesized skill profiles and conducting intervention studies to identify appropriate remediation for each profile. Each is discussed in some detail next.

To further validate and refine the reading competency profiles presented here, we attempt to model the test performance specific to each profile using a new psychometric approach called the *attribute hierarchy method* (AHM; Gierl, Leighton, & Hunka, 2000; Leighton, Gierl, & Hunka, 2004). The AHM is a psychometric method for classifying examinees' item responses into a set of structured attribute patterns associated with different components from a cognitive model of task performance (Leighton & Gierl, 2006). An *attribute* is a description of the procedural or declarative knowledge needed to answer an item successfully. The examinee must possess and exercise these attributes sequentially according to the proposed hierarchy to answer items correctly.

To generate the attribute hierarchy for each profile, the SAT critical reading items will be coded according to a set of ideal solution strategies that reflect how examinees in each reading profile are expected to solve the items. The coding task will entail evaluating each item to determine what component processes would primarily be measured for an examinee within each profile. What will be the probability of success for a Profile 3 examinee (weak lower level, strong higher level processing skills), for example, on an item coded to measure primarily lower level processes (Word Meaning and Sentence Meaning)? Then, using the AHM, the sequence of attributes measured by the items, as specified in the hierarchy, will be applied to the response patterns to model the likely solution path for examinees in each profile. We assume that the processes are sequential and dependent, as examinees must invoke each attribute, which depend on prior attributes, to generate a correct solution. We also hypothesize that higher level processes can be called on to compensate for deficits in some lower level processes (if a sufficient threshold of information is generated through the lower level processes to support higher level processing) (McNamara, Kintsch, Songer, & Kintsch, 1996; Perfetti, 1985a). Items that lend themselves to such compensatory processing would produce a different ideal response pattern for Profile 3 examinees, for instance, than items for which higher level processes would not compensate for deficits in lower level processes. The probability of success on an item will be calculated for examinees in each profile as a function of the probabilities of the sequential and dependent attributes modeled for the items associated with each profile. Verbal report data will also be collected and mapped onto the outcomes from the AHM analyses as a way of validating our cognitive inferences about the examinees in each profile.

A second method that we will use to refine the hypothesized skill model is item difficulty modeling. Preliminary research by The College Board (Huff, 2006) indicates that when critical reading items are grouped into mutually exclusive categories according to the dominant skill categories (i.e., Word Meaning, Sentence Meaning, Situation Model, Global Text Meaning, or Pragmatic Meaning), there is no relationship to item difficulty. This result was expected given that during the task analysis, raters focused on item attributes that would be applicable outside the assessment context. However, preliminary results indicate that when items are coded to more than one skill category, including a measure of the degree to which each skill is required to solve the item (high, medium, low), then the predictive power of the text processing skills increases greatly. Future research will explore why coding items to more than one skill category improves the relationship between the skill categories and item difficulty. We will also investigate how this relationship between skill categories and item difficulty can be used to refine the hypothesized profiles; relate the profiles to total test score; and, potentially, inform test design.

Once examinees can be reliably and validly classified into one of the four profiles, remediation techniques that are specific to each profile should be developed. One approach to identifying and developing remediation is to design and conduct *dimension–treatment interactions*, such as through a repeated-measures experimental design. For example, pretests establish an individual examinee's reading ability profile; treatments targeting either lower level linguistic processing skills or higher level text modeling skills will be provided, based on the profile; posttests will produce contrasting subscores on the component processes and the summary critical reading score. By regressing the posttest subscores and summary score onto the pretest scores, the treatment interaction can be determined. The relative contributions of treatments targeting lower level and higher level processes to the summary score can be contrasted, and inferences can be drawn about which component processes are most efficient in their contribution to reading ability. By contrasting the contributions of both lower level and higher level processes to reading ability, these analyses will shed further light on Perfetti's conclusion that lower level linguistic processes associated with lexical access and proposition assembly account for the greatest differences in reading ability. Results from these investigations should also inform the selection of treatments for examinees fitting Profiles 2, 3, and 4, depending on the relative efficiency of the contrasted dimensions for improving overall reading ability.

We find the results from these preliminary studies promising for using the SAT critical reading section to reliably categorize students into reading competency profiles that support matching students to effective reading comprehension treatments. Our task analyses and statistical data modeling have validated that the multiple text processing skills required by each item can be parsed out and measured through careful and purposeful coding. Because no prior research has been conducted to investigate explicitly whether the SAT measures multiple complementary dimensions that could support differentiated instructional decisions, these findings are both unprecedented and promising.

Although much work remains, the potential benefits of these results are substantial. By parsing, measuring, and reporting on the sophisticated, interdependent text processing skills that enable students to learn from texts in the content areas, we can inform how others design systematic reading comprehension instruction in the middle and secondary grades. As these research goals come to fruition, the SAT critical reading section can be one piece of a coherent curriculum, instruction, and assessment system designed to help the 8 million struggling readers in Grades 4 through 12 (National Center for Education Statistics, 2003) graduate high school with the critical reading skills they will need to succeed in college and the 21st-century workplace.

REFERENCES

Anderson, R. C. (1982). How to construct achievement tests to assess comprehension. *Review of Educational Research*, *42*, 145–170.

Bolt, D. M. (2005). Limited- and full-information estimation of item response theory models. In A Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics* (pp. 27–71). Mahwah, NJ: Lawrence Erlbaum Associates.

Buck, G., VanEssen, T., Tatsuoka, K., Kostin, I., Lutz, D., & Phelps, M. (1998). *Development, selection and validation of a set of cognitive and linguistic attributes for the SAT I Verbal: Sentence completion section*. Educational Testing Service Research Report RR-98-23, Princeton, NJ.

Carnegie Council on Adolescent Development. (1989). *Turning points: Preparing American youth for the 21st century, report of the Task Force on Education of Youth Adolescents*. New York: Carnegie Corporation of New York.

Cook, L. J., Dorans, N. J., & Eignor, D. R. (1988). An assessment of the dimensionality of three SAT-Verbal test editions. *Journal of Educational Statistics*, *13*, 19–43.

Daneman, M., & Carpenter, R. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

DiBello, L. V. (2002, April). *Skills-based scoring models for the PSAT/NMSQT*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

DiBello, L. V., & Crone, C. (2001, April). *Technical methods underlying the PSAT/NMSQT™ enhanced score report*. Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, WA.

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361–390). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Dorans, N., & Lawrence, I. (1999). *The role of the unit of analysis in dimensionality assessment*. Educational Testing Service Research Report RR-99-14, Princeton, NJ.

Embretson, S. E., & Wetzel, C. D. (1987). Component latent trait models for paragraph comprehension. *Applied Psychological Measurement*, *11*, 175–193.

Everson, H. (2004). Innovation and change in the SAT: A design framework for future college admission tests. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 75–91). New York: Routledge-Falmer.

Frederiksen, J. R. (1982). A componential theory of reading skills and their interactions. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence: Vol. 1* (pp. 125–180). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Froelich, A. G. (2001, April). *Assessing the unidimensionality of test items and some asymptotics of parametric item response theory*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Froelich, A. G., & Habing, B. (2001). *Refinements of the DIMTEST methodology for testing unidimensionality and local independence*. Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, WA.

Gierl, M. J. (2004). *Identifying cognitive dimensions that affect student performance on the New SAT: A discussion paper*. Paper presented to the College Board, New York.

Gierl, M. J., Bisanz, J., Bisanz, G., & Boughton, K. (2003). Identifying content and cognitive skills that produce gender differences in mathematics: A demonstration of the DIF analysis framework. *Journal of Educational Measurement*, *40*, 281–306.

Gierl, M. J., Leighton, J. P., & Hunka, S. M. (2000). Exploring the logic of Tatsuoka's rule-space model for test development and analysis. An NCME Instructional Module. *Educational Measurement: Issues and Practice*, *19*(3), 34–44.

Gierl, M. J., Leighton, J. P., Wang, C., & Tan, X. (2005). *Identifying cognitive dimensions that affect student performance on the New SAT—Technical Report No. 2: Evaluating primary skill categories*. New York: College Examination Board.

Gierl, M. J., Tan, X., & Wang, C. (2005). *Identifying content and cognitive dimensions on the SAT* (Research Report No. 2005-11). New York: College Examination Board.

Glaser, R. (1986). A cognitive science perspective on selection and classification and on technical training. *Advances in Reading/Language Research*, *4*, 253–268.

Gorin, J. S. (2005). Manipulating processing difficulty of reading comprehension questions: The feasibility of verbal item generations. *Journal of Educational Measurement, 4,* 351–373.

Greeno, J. G. (1976). Cognitive objectives in instruction: Theory of knowledge for solving problems and answering questions. In D. Klahr (Ed.), *Cognition and instruction* (pp. 123–159). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Greeno, J. G. (1989). A perspective on thinking. *American Psychologist*, *44*, 134–141.

Griffin, P. (2001, April). *Performance assessment of higher order thinking*. Paper presented at the annual conference of the American Education Research Association, Seattle, WA.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component process of reading comprehension. *Journal of Educational Psychology*, *93*, 103–128.

Huff, K. (2004, April). A practical application of evidence centered design principles: Coding items for skills. In K. Huff (Organizer), *Connecting curriculum and assessment through meaningful score reports*. Symposium conducted at the meeting of the National Council on Measurement in Education, San Diego, CA.

Huff, K. (2006, April). *Using item difficulty modeling to inform descriptive score reports*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Huff, K., & Goodman, D. (2007). The demand for cognitive diagnostic assessment. In J. Leighton & M. Gierl (Eds.), *Cognitively diagnostic assessment*. London: Cambridge University Press.

Hunt, E. (1978). Mechanics of verbal ability. *Psychological Review*, *85*, 109–130.

Hunt, E. (1985). Verbal ability. In R. J. Sternberg (Ed.), *Human abilities: An information-processing approach* (pp. 31–58). New York: Freeman.

Hunt, E., Frost, N., & Lunnenborg, C. (1973). Individual differences in cognition: A new approach to intelligence. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7, pp. 87–122). New York: Academic.

Hunt, E., & Lansman, M. (1975). Cognitive theory applied to individual differences. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Introduction to concepts and issues* (Vol. 1, pp. 81–107). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hunt, E., Lunnenborg, C., & Lewis, J. (1975). What does it mean to be high verbal? *Cognitive Psychology*, *7*, 194–227.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *98*, 122–149.

Kail, R. V., Chi, M. T. H., Ingram, A. L., & Danner, F. W. (1977). Constructive aspects of children's reading comprehension. *Child Development*, *48*, 684–688.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, MA: Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394.

Kirsch, I. S., & Mosenthal, P. B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly*, *25*, 5–30.

Leighton, J. P., Gierl, M. J., & Hunka, S. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, *41*, 205–237.

Leighton, J. P., & Gierl, M. J. (2006). *Identifying and evaluating cognitive models in educational measurement*. Manuscript submitted for publication.

Leighton, J. P., & Gierl, M. J. (in press). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*.

Lesgold, A. M., & Perfetti, C. A. (1978). Interactive processes in reading comprehension. *Discourse Processes*, *1*, 323–336.

Linn, R. L. (1986). Educational testing and assessment: Research needs and policy issues. *American Psychologist*, *41*, 1153–1160.

Liu, J., Feigenbaum, M., & Walker, M. E. (2004, April). *New SAT and New PSAT/NMSQT Spring 2003 field trial design*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Louwerse, M., & van Peer, W. (2002). *Thematics: Interdisciplinary studies*. Amsterdam/Philadelphia: John Benjamins.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 357–407.

McNamara, D., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education, Macmillan.

Mislevy, R. J. (1994) Evidence and inference in educational assessment. *Psychometrika*, *59*, 439–483.

Mislevy, R. J. (1996). *Evidence and inference in educational assessment* (CSE Technical Report 414). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education and Information Studies, University of California.

Mislevy, R. J., & Riconscente, M. M. (2005). Evidence-centered assessment design. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 61–90). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mosenthal, P. B., & Kirsch, I. S. (1991). Toward an explanatory model of document literacy. *Discourse Processes*, *14*, 147–180.

National Center for Education Statistics. (2003). *Nation's report card: Reading 2002*. Washington, DC: U.S. Government Printing Office.

National Education Goals Panel. (1991). *Measuring progress toward the national educational goals: Potential indicators and measurement strategies*. Washington, DC: U.S. Government Printing Office.

Nichols, P. D. (1994). A framework for developing cognitively diagnostic assessments. *Review of Educational Research*, *64*, 575–603.

Nitko, A. J. (1989). Designing tests that are integrated with instruction. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 447–474). New York: American Council on Education, Macmillan.

O'Neill, K. A., & McPeek, W. M. (1993). Item and test characteristics that are associated with differential item functioning. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 255–279). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Pellegrino, J. W., Baxter, G. P., & Glaser, R. (1999). Addressing the "two disciplines" problem: Linking theories of cognition and learning with assessment and instructional practice. *Review of Research in Education*, *24*, 307–353.

Pellegrino, J. W., & Glaser, R. (1979). Cognitive correlates and components in the analysis of individual differences. *Intelligence*, *3*, 187–214.

Pellegrino, J. W., Jones, L. R., & Mitchell, K. J. (Eds.). (1999). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress*. Washington, DC: National Academy Press.

Perfetti, C. A. (1985a). *Reading ability*. New York: Oxford University Press.

Perfetti, C. A. (1985b). Reading ability. In R. J. Sternberg (Ed.), *Human abilities: An information-processing approach* (pp. 31–58). New York: Freeman.

Perfetti, C. A., & Goldman, S. R. (1976). Discourse memory and reading comprehension skill. *Journal of Verbal Learning and Verbal Behavior*, *15*, 33–42.

Perfetti, C. A., & Hoagaboam, T. (1975). Relationship between simple word decoding and reading comprehension skill. *Journal of Educational Psychology*, *67*, 461–469.

Schum, D. A. (1987). *Evidence and inference for the intelligence analyst*. Lanham, MD: University Press of America.

Shealy, R., & Stout, W. F. (1993). A model-based standardization approach that separates true bias/DIF from group differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika*, *58*, 159–194.

Sheehan, K. M., & Ginther, A. (2001, April). *What do passage-based multiple-choice verbal reasoning items really measure? An analysis of the cognitive skills underlying performance on the current TOEFL reading section*. Paper presented at the 2001 annual meeting of the National Council of Measurement in Education, Seattle, WA.

Snow, R. E. (1989). Aptitude interactions as a framework for individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences: Advances in theory and research* (pp. 13–59). New York: Freeman.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 263–331). New York: American Council on Education, Macmillan.

Sternberg, R. J. (1977). Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Stiggins, R. (1994). *Student-centered classroom assessment*. New York: Merrill.

Stout, W., Habing, B., Douglas, J., Kim, H. R., Roussos, L., & Zhang, J. (1966). Conditional covariance-based nonparametric multidimensional assessment. *Applied Psychological Measurement, 20,* 331–354.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*, 345–354.

Tatsuoka, K. K. (1984). Caution indices based on item response theory. *Psychometrika*, *49*, 95–110.

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics*, *10*, 55–73.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnoses. In N. Frederiksen, R. L. Glaser, A. M. Lesgold, & M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453–488). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Tatsuoka, K. K. (1991). *Boolean algebra applied to determination of universal set of knowledge states*. Research Report No. RR-91-44-ONR, Educational Testing Service, Princeton, NJ.

Tatsuoka, K. K., & Tatsuoka, M. M. (1991). On measures of misconception stability. Research report, ONR Grant N00014-86-K0290, University of Illinois, Urbana.

Tatsuoka, M. M., & Tatsuoka, K. K. (1987). Rule space. In S. Kotz & N. Johnson (Eds.), *Encyclopedia of statistical sciences* (pp. 217–221). New York: Wiley.

VanderVeen, A. (2004, April). Toward a construct of critical reading for the new SAT. In K. Huff (Organizer), *Connecting curriculum and assessment through meaningful score reports*. Symposium conducted at the meeting of the National Council on Measurement in Education, San Diego, CA.

Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General*, *117*, 319–331.

III

Reading Comprehension Strategy Interventions

7

Increasing Strategic Reading Comprehension With Peer-Assisted Learning Activities

Douglas Fuchs and Lynn S. Fuchs
Vanderbilt University

The authors describe a multiyear research and development program that has produced peer-assisted learning strategies (PALS) in reading. They describe this program of research and relevant findings and explain the PALS procedures for children from preschool through the intermediate elementary grades. They discuss 5 conclusions from this work. First, PALS is a promising platform for promoting strategic reading comprehension behavior; that is, using the PALS structure and organization, they have transformed research-based instructional procedures, developed in laboratory-like settings, into feasible practices that can be implemented by real teachers in real schools. Second, young children can be taught to work in pairs, using structured activities, so that they follow procedures accurately, with productive outcomes. Third, they have explored and found no interaction between student type (disabled, low, average, high) and treatment condition, or between Title I and middle-class schools and treatment. This suggests that the potential applicability of PALS across a wide range of settings for enhancing strategic reading comprehension behavior. Fourth, despite the apparent robustness of PALS, they have across their studies identified nonresponders; that is, children who have not benefited from PALS implementation. For this reason, they argue, responsible educational practice demands ongoing monitoring of student *progress so that nonresponders can be identified promptly and helped appropriately. Finally, the authors' foray into 1st-grade reading comprehension was not successful. Students who received instruction limited to word reading skill outperformed those whose PALS reading time was continually interrupted with activities designed to promote strategic reading comprehension skill. This suggests that 1st-grade reading comprehension strategy instruction may be unproductive when those children have underdeveloped word-level skills.*

Today's public school classrooms represent ever-increasing profiles of diversity. Societal forces, such as immigration (cf. Hodgkinson, 1995), lead to linguistic, ethnic, and cultural differences, which afford many important advantages. Another form of classroom diversity, however, challenges the capacity of schools to provide appropriate instruction. That is, education policies such as "detracking" (e.g., Braddock et al., 1992) and "inclusion" (D. Fuchs & Fuchs, 1994) have converged to make public school classrooms more academically diverse. In fact, the degree of academic heterogeneity in the typical urban classroom is striking. For example, in recent work in the area of reading (L. S. Fuchs & Fuchs, 2003) conducted in urban second-grade classrooms, teachers averaged 22 children whose curriculum-based measurement reading scores (i.e., words read correctly in 1 minute from connected text) within a single classroom ranged, on average, between 0 and 183.

This heterogeneity strains the capacity of teachers to address students' needs in the area of reading, and many teachers respond to this diversity in reading skill by ignoring it. That is, most reading lessons are designed for, and directed to, the group of students performing at the middle of the class (e.g., Baker & Zigmond, 1990; D. Fuchs, Fuchs, Mathes, & Simmons, 1997; L. S. Fuchs, Fuchs, Phillips, & Simmons, 1993). The needs of other children often go unmet. This may help explain why so many children read poorly and eventually are labeled learning disabled.

Thus, one critical question facing public school instruction today is how to differentiate reading instruction to address growing academic heterogeneity. Over the last 20 years, a strategy that has gained increasing popularity is to supplement conventional teaching with collaborative, peer-mediated instruction, whereby children work together to support each other's reading development. Peer-mediated instruction capitalizes on such academic diversity by creating subgroups of children in the same classroom who operate on different levels of curricula and use different instructional procedures. That way, teachers can conduct many simultaneous lessons and address a broader range of students' educational needs. Research in the elementary grades shows that children's reading competence improves when they work with each other in a cooperative and structured manner (e.g., Greenwood, Delquadri, & Hall, 1989; Rosenshine & Meister, 1994; Stevens, Madden, Slavin, & Farnish, 1987).

Two collaborative learning methods in the area of reading for which a strong research base exists are *Cooperative Integrated Reading and Composition* (CIRC; Stevens et al., 1987) and *Reciprocal Teaching* (RT; Palincsar & Brown, 1984). CIRC teachers conduct daily small-group basal instruction along with weekly direct instruction reading lessons in comprehension-fostering and metacognitive strategies, using special CIRC materials. Students work in mixed-ability learning teams to master the basal and direct instruction lesson content, using stimulus materials related to the basal text, which teams of teachers develop. Collaborative activities include oral reading in pairs, decoding, working on story structure, prediction, and story summary activities. A cooperative reward structure is used, whereby groups are responsible for individual student learning, and teams receive certificates and other rewards.

With RT, students read a passage of expository material, paragraph by paragraph. While reading, they learn and practice how to generate questions, summarize, clarify word meanings and confusing text, and predict subsequent paragraphs. In the early stages of RT, the teacher models these strategies, and then students practice the strategies on the next section of the text as the teacher tailors feedback through modeling, coaching, hints, and explanations. The teacher also invites students to react to peers' statements by elaborating or commenting, suggesting other questions, requesting clarifications, and helping to resolve misunderstandings. In the course of this guided practice, the teacher gradually shifts responsibility to the students for mediating discussions, as the teacher observes and helps as needed. At this point, sessions become dialogues among students as they support each other and alternate among prompting the use of a strategy, applying and verbalizing that strategy, and commenting on the application.

These programs are well developed and tested. Nevertheless, in our work in the schools, we were reluctant to adopt either approach because of feasibility and usability concerns. With respect to CIRC, a substantial commitment is demanded from teachers in terms of the percentage of their allocated reading instruction time. Also, CIRC requires schools to invest considerable effort in creating and duplicating materials. RT presents a different set of feasibility and usability concerns. First, RT may be inappropriate for younger elementary age children, where effects are unclear (Rosenshine & Meister, 1994). Second, RT can be difficult for teachers to master because the strategic comprehension behaviors are unfamiliar to many teachers (Pressley, 1997) and because the techniques for helping children develop responsibility for and competence in engaging peers' strategic behavior are challenging.

Given these feasibility and usability issues, we initially were attracted to a third validated form of collaborative learning, *Class Wide Peer Tutoring* (CWPT; Greenwood et al., 1989). CWPT is easier for schools to adopt and use for three reasons. First, it requires no materials development and only minimal duplication at the beginning of the school year. Second, it occurs only three times weekly, for 40 minutes per session, and is designed to supplement any instructional approach the teacher uses. Last, it is simple for teachers and students to learn. With CWPT, children work in pairs; one child reads aloud for 5 minutes while the partner identifies and corrects errors; the partner asks who-, what-, why-, where-, and when-questions for 5 minutes; then students switch roles and repeat activities.

At the same time, however, we questioned whether the CWPT reading activities could be enriched, and effectiveness thereby enhanced, without detracting from implementation ease. So, in 1989 we set out to develop and research a set of collaborative learning methods that would be appropriate for children throughout the primary and intermediate grades, incorporate the richness represented in CIRC and RT, and benefit from CWPT's feasibility. With this goal in mind, we borrowed the CWPT structure, but we substituted a potentially richer set of activities. Our first experiment tested CWPT against this modified and enriched CWPT. Since 1989, we have built and researched a modified CWPT structure in reading, known as *Peer-Assisted Learning Strategies* (PALS).

In this chapter, we summarize the major components of this research program, which focuses primarily on comprehension skill at Grades 2 through 6. At preschool, kindergarten, and first grade, PALS activities are designed to also help children develop word-level reading skills, but as we describe first-grade work in this chapter, we primarily address reading fluency and comprehension. Toward that end, we highlight four investigations that illustrate three types of studies in the PALS program of research. The first study represents our early work in which, as we developed PALS, we contrasted the contribution of alternative components. The second and third studies are experimental evaluations of PALS's overall effectiveness, conducted once major components of the treatment had been specified. One of these overall comparisons was conducted with children whose primary language was English; the second assessed effects on English language learners. In the last study, we describe a more recent extension to the PALS research program at first grade, where we explicitly targeted strategic reading comprehension.

CONTRIBUTION OF ALTERNATIVE COMPONENTS

In the first major PALS experiment, conducted in 1990–1991 (Simmons, Fuchs, Fuchs, Pate, & Mathes, 1994), we asked two questions: (a) Does an enriched, but more complicated, set of peer-mediated activities support greater student learning than does the more simple set of CWPT activities? and (b) Does role reciprocity, whereby both students serve as tutor and tutee in each session, enhance learning?

Rationale for the Study Foci

CWPT comprises two primary components: (a) organizational procedures that provide structure for scheduling, pairing students, arranging the physical environment, and monitoring student participation, and (b) content-specific activities (i.e., sustained reading and answering who-, what-, where-, when-, and why-questions). CWPT's organizational structure is elegant; we wished to maintain it. As already discussed, however, we were less certain about the CWPT activities. Although the simplicity of the CWPT activities has merit within increasingly demanding instructional contexts, a number of approaches promote reading fluency and comprehension in more strategic, although more complex, ways. Given the potential efficacy of these more complex approaches over those of CWPT, we decided to contrast CWPT activities to more complex activities while maintaining the CWPT's organizational structure. We incorporated three enriched and more complex activities (which we substituted for CWPT's sustained reading with who-, what-, why-, where-, and when-questions): repeated reading, paragraph summaries, and prediction activities.

Posner and Synder's (1975a, 197b) theory of expectancy provides a framework for conceptualizing repeated reading as a method for enhancing reading comprehension. The theory poses that semantic context affects word recognition by means of two independently acting processes. With the *automatic-activation process*, stimulus information activates a memory location and spreads automatically to semantically related memory locations that are nearby in the network. This process is obligatory, fast acting, and requires no attentional capacity. The second process, a *conscious-attention mechanism*, relies on context to formulate a prediction about the upcoming word and directs the limited-capacity processor to the memory location of the expected stimulus. This slow-acting process is optional, it uses attentional capacity, and it inhibits the retrieval of information from unexpected locations. For good readers, rapid word recognition short circuits the conscious-attention mechanism; the automatic spreading-activation components of contextual processing dominate. By contrast, for poor readers, contextual facilitation results from the combined effect of the conscious-attention and the automatic-activation mechanisms. Unfortunately, as poor readers rely on the conscious-attention mechanism, they expend their capacity in prediction processes to aid in word recognition, and little is left over for integrative comprehension processes. The goal of repeated reading is to help readers transition to the automatic-activation mechanisms to free up capacity for higher level, integrative comprehension processing of text. Repeated reading is viewed as one instructional procedure for achieving this transition.

In fact, in previous work, repeated reading has increased words read correctly per minute, decreased word recognition errors, and improved comprehension (Herman, 1985; O'Shea, Sindelar, & O'Shea, 1987; Samuels, 1979). With peer mediation, repeated reading was conducted as follows. The reader read a passage for 1 minute (the teacher signaled the beginning and end of the minute for the whole class) while the peer tutor recorded errors. The tutor corrected and provided feedback on errors, using a standard correction procedure, and then tallied 1 point for each correctly read sentence. This sequence recurred twice, so that both readers reread the same text three times.

The second activity, paragraph summary, was a modification of the paragraph restatement strategy developed and tested by Jenkins, Heliotis, Haynes, and Beck (1986). Consistent with a constructionist model proposed by Graesser, Singer, and Trabasso (1994), the strategy encourages readers to construct meaning representations that are coherent at the local level (and to connect these representations at a more global level with the prediction strategy, which we describe next). Yet, research shows that generating summaries can be difficult. That is, students cannot adequately summarize typical fifth-grade material until well into high school, and junior college poor readers have not yet mastered this activity (Brown & Day, 1983). Paragraph summarization requires readers to monitor comprehension and make conscious judgments in the selection and reduction of textual information (Hidi & Anderson, 1986; chap. 3, this volume); allocate attention to the major content and check if they have understood it (chap. 3, this volume; Palincsar & Brown, 1984); and, as an application of the generative process model of reading, to elaborate on the information provided in text (Doctorow, Wittrock, & Marks, 1978). In these ways, paragraph summary reflects cognitive and metacognitive processes, and overt practice in generating summaries should promote those processes. In light of the difficulties children experience, primary and intermediate grade children might benefit from practice in formulating summaries. In fact, research shows that practice in paragraph summaries, which require identification of main ideas, enhance reading comprehension (Baumann, 1984; Bean & Steenwyk, 1984; Paris, Cross, & Lipson, 1984; Rinehart, Stahl, & Erickson, 1986). Within a peer-mediated situation like PALS, paragraph summary occurred as follows. The first reader read a paragraph aloud while the tutor identified and corrected errors. Then, the reader stopped to identify who or what the paragraph was mostly about and the most important thing that happened. The tutor prompted corrections as necessary.

The third activity, prediction relay, extends paragraph summary to half-pages of text and requires formulation and checking of predictions. As already noted, this strategy, which is consistent with a constructionist model, encourages readers to tie summaries together at a more global level. Prediction is a coherence and explanation strategy (see chap. 1, this volume), which requires inference modeling, as described by Oakhill and Cain (chap. 3, this volume) As Palincsar and Brown (1984) described, expert readers proceed automatically through text until an event alerts them to a comprehension failure, which then prompts debugging activities. One common triggering event is the realization that an expectation about the text has not been confirmed. This premise assumes that expert readers automatically formulate and check predictions as they read. Research, however, documents that young and poor readers have difficulty evaluating text for internal consistency and compatibility with known facts (Englert & Hiebert, 1984; Markman, 1981) and that the ability to interpret what will occur next in text develops slowly (Collins & Smith, 1982). With peer mediation, in this study, prediction relay was designed to help children develop and automatize the strategic behavior of formulating and checking predictions about the text they read, by practicing that strategy in an overt manner. With prediction relay, the reader made a prediction about the upcoming half-page, which the tutor either acknowledged as plausible or requested a correction; read that text aloud while the tutor identified and made corrections; (dis)confirmed the prediction while the tutor checked and prompted corrections; and proceeded on to subsequent half-pages in the same way.

Beyond the issue of instructional complexity, our second question focused on role reciprocity. According to role theory, academic gains effected by peer and crossage tutoring may be attributable to the enactment of a role that produces changes in behavior, attitudes, and self-perceptions (Allen, 1976). Most peer tutoring occurs with stronger students acting as tutors and weaker students as tutees, even though some research suggests the potential for reciprocal tutoring (Top & Osguthorpe, 1987; Wiegmann, Dansereau, & Patterson, 1992). Moreover, as a generative model of learning (Wittrock, 1989) indicates, a tutorial role requires students to engage in active monitoring to identify and correct errors and to elaborate on information in their explanations, activities that may benefit lower performing students.

Method and Results

The study participants were 31 general educators in Grades 2 through 5. In these classes, where treatments were implemented classwide, we identified 118 children as target research participants. They were in the lower performing half of each class; 58 had been identified as having a learning disability, 27 were low performers never referred to special education, and 33 performed in the average range. Twenty-three classrooms were assigned randomly to CWPT or to more complex collaborative activities (i.e., repeated readings, paragraph summaries, and prediction relay). Half the classrooms in each condition were assigned randomly to a role reciprocity condition. Eight teachers and their targeted students constituted the contrast condition, in which the same amount of total reading instruction occurred without collaborative learning. Tutoring took place three times each week for 14 weeks.

Immediately before and after treatment, we administered the Comprehensive Reading Assessment Battery (CRAB; L. S. Fuchs, Fuchs, & Hamlett, 1989) to each target child. Using four folk tales (Brown & Smiley, 1977), two at each administration, the CRAB produces five scores: (a) average number of words read aloud correctly in 3 minutes, (b) average number of questions answered correctly, (c) number of words and (d) number of content words written on a recall of the passage, and (e) number of maze (or multiple-choice cloze) items restored correctly. Treatment fidelity, measured four times in each classroom with direct observation, revealed strong implementation across conditions and time, with the mean percentage of correctly implemented components for the four tutoring conditions ranging from 89.78 to 94.54. We analyzed the data using classroom as the unit of analysis and student type as a within-classroom variable. We found no significant interactions between treatment and type of students, indicating that effects applied comparably across learning disabled, low-performing, and average-performing children.

For reading fluency, outcomes revealed the following. Students in all tutoring conditions made significantly greater growth than those in the contrast condition, and effect sizes ranged from 0.29 to 0.41 *SD*. There were, however, no significant differences between tutoring groups. One explanation for this lack of difference is that all tutoring conditions provided considerable opportunity for students to read and receive corrections in a carefully structured manner. Therefore, potential differences among conditions may have been obscured by a more potent commonality: increased opportunity to read and receive systematic feedback. In addition, the lack of difference on reading fluency between more and less complex activities may be attributable to the nature of our repeated reading treatment. The efficacy of repeated reading has been documented in short-term treatments (O'Shea et al., 1987) with controlled passages that maximize common words (Samuels, 1979). By contrast, our repeated reading treatment lasted a relatively long time and used uncontrolled basal text or library material.

Also on reading fluency, we found no differences between students who served in one or both tutoring roles. According to role theory (Bierman &

Furman, 1981; Cohen, 1984), the opportunity to act as tutor may increase self-perceptions of competence enough to improve academic performance. Yet the relation between academic learning time and achievement suggests the benefits of increased opportunity to read, which was afforded students who participated only as tutees. Moreover, although Wittrock's (1989) model of generative learning postulates that constructing explanations may enhance children's own understanding, the role of tutoring during reading aloud provides few opportunities for tutors to elaborate. In any case, the absence of differences between role conditions, when combined with findings of greater learning for both instructional complexity conditions over that of a contrast group, indicated that reading fluency was enhanced with either configuration. Given the complexity associated with the repeated reading condition, we opted to design PALS with sustained, rather than repeated, reading.

What about comprehension? Only students who participated in the reciprocal, complex activities improved more than contrast students on comprehension measures; the average effect size was 0.65. With respect to reciprocity, we speculated that, in contrast to the reading-aloud activities, paragraph summary and prediction activities did afford children greater opportunity to benefit from the tutoring role. By means of the status associated with the tutoring role or by means of the actual metacognitive and cognitive demands associated with that role (i.e., monitoring partners' construction of main idea statements and checking the plausibility and actuality of partners' predictions), lower performing children made greater gains than comparable students who only had opportunities to serve as tutees. This finding, which lends support to role theory and a generative model of learning, led us to adopt role reciprocity within PALS. Comprehension findings also supported the efficacy of the more complex tutoring activities, at least in combination with role reciprocity. As discussed, prior work supports overt practice in strategic comprehension activities such as paragraph summaries (Baumann, 1984; Bean & Steenwyk, 1984; Hare & Borchardt, 1984; Jenkins et al., 1986; Rinehart et al., 1986) and prediction activities (e.g., Rosenshine & Meister, 1994). We therefore decided to incorporate the summary and prediction activities into PALS.

ESTIMATING OVERALL TREATMENT EFFECTS

Over several years, we have cumulatively tested the contribution of contrasting components and finally arrived at a PALS treatment for which the component parts appeared effective and which teachers found workable in typical classrooms. In the next phase of our research program, therefore, we assessed the overall efficacy of the treatment. We describe the PALS treatment tested in that series of studies, then we illustrate this phase with two investigations, one conducted with children whose primary language was English (D. Fuchs, Fuchs, Mathes, & Simmons, 1997) and the other conducted with children with limited English proficiency (Saenz, Fuchs, & Fuchs, 2005).

The PALS Treatment

Each week, teachers incorporate three 35-minute PALS sessions into their allocated reading time, implementing PALS with all children in their classes. Teachers begin by conducting seven lessons on how to implement PALS. Each lesson lasts 45 to 60 minutes and incorporates teacher presentations, student recitation of information and application of principles, and teacher feedback on student implementation (see D. Fuchs, Fuchs, Mathes, & Simmons, 1996, for the manual).

During PALS, every student in the class is paired; each pair includes a higher and lower performing student. The teacher determines dyads by ranking the class on reading competence, doing a median split, and pairing the highest performer from the top half with the highest performer from the bottom half, and so on. Although tutoring roles are reciprocal, the higher performing student reads first for each activity to serve as model for the lower performing student. Both students read from material appropriate for the lower reader, which typically is literature the teacher selects at the appropriate difficulty level. Pairs are assigned to one of two teams for which they earn points. Points are awarded for completing reading activities correctly and demonstrating good tutoring. Each pair keeps track of points on a consecutively numbered scorecard, which represents joint effort and achievement. Each time a student earns a point, the tutor slashes the next number. Also, as teachers lead PALS sessions, they circulate and award points to reward cooperative behavior and correct tutoring methods. At the end of the week, each pair reports the last number slashed as the pair's total; the teacher sums each team's points, and the class applauds the winning team. Every 4 weeks, the teacher formulates new pairs and team assignments. Thus, the motivational system combines competitive (team vs. team) and cooperative (combined effort of the pair) structures.

The first activity in every PALS session is Partner Reading. Each student reads connected text aloud for 5 minutes, for a total of 10 minutes. The higher performing student reads first; the lower performing student rereads the same material. After both students have read, the lower performing student retells for 2 minutes the sequence of what occurred. Students earn 1 point for each correctly read sentence and 10 points for the retell.

The second PALS activity, Paragraph Shrinking, is designed to develop comprehension through summarization and main idea identification. Continuing to read subsequent sections of text, students read orally one paragraph at a time, stopping to identify its main idea. Tutors guide the identification of the main idea by asking readers to identify who or what the paragraph is mainly about and the most important thing about the who or what. Readers put these two pieces of information together in 10 or fewer words. For each summary, students earn 1 point for correctly identifying who or what, 1 point for correctly stating the most important thing about the who or what, and 1 point for using no more than 10 words. Students continue to monitor and correct reading errors, but points no longer are awarded for reading sentences correctly. After 5 minutes, students switch roles.

The last activity is Prediction Relay. It extends Paragraph Shrinking to larger chunks of text and requires students to formulate and check predictions. Prediction Relay comprises five steps. The reader makes a prediction about what will be learned on the next half-page; reads the half-page aloud while the tutor corrects errors, (dis)confirms the prediction, and summarizes the main idea. Students earn 1 point for each viable prediction, 1 point for reading each half-page, 1 point for accurately confirming each prediction, and 1 point for each summary component (the who or what, what mainly happened, and 10 or fewer words). After 5 minutes, students switch roles.

Effects on English-Proficient Students of Varying Achievement Status

To study effects on English-proficient students of varying achievement status, we (Fuchs et al., 1977a) assigned 12 schools, stratified on achievement and family income, to experimental and control groups. At Grades 2 through 6, 20 teachers implemented PALS; 20 did not. PALS teachers implemented the treatment classwide, but we identified three students in each class as research participants: (a) one with an identified learning disability in reading (LD), (b) one low achiever never referred for special education (LA), and (c) one average achiever (AA). All students had English as their primary language. We tested each research participant with the CRAB before and after treatment, which lasted 15 weeks. Fidelity data, collected three times with observation, documented strong implementation. Instructional plan sheets revealed that PALS and no-PALS teachers allocated comparable time to reading instruction.

We analyzed the achievement data using treatment, trial, and student type (LD, LA, AA) as factors (classroom was the unit of analysis; trial was a repeated measure; student type was a within-classroom variable). We found significant Treatment × Trial interactions on all CRAB scores, indicating that, compared with conventional instruction, PALS students grew more on fluency, accuracy, and comprehension. Moreover, the three-way interaction was not statistically significant, so treatment effects were not mediated by the students' initial achievement status. Aggregated across the three types of students, effect sizes were 0.22, 0.55, and 0.56 on the CRAB words read correctly, questions answered correctly, and maze blanks restored correctly, respectively. These effects compare favorably with more complicated versions of collaborative learning. As Slavin (1994) reported, the median effect size for 52 studies of cooperative learning treatments that lasted more than 4 weeks was .32, a figure identical to the one reported by Rosenshine and Meister (1994) for RT.

Effects on Students With Limited English Proficiency of Varying Achievement Status

The study assessing effects on limited English proficiency (LEP) students (Saenz et al., 2005) paralleled the Fuchs et al. (1997a) investigation just described, with these important differences. First, the participants were 12 teachers in south Texas, working in schools that served LEP populations. From each class, we sampled only students (n = 132) who were native Spanish-speaking and identified by their school district as LEP according to Texas state eligibility criteria. Second, the study was conducted at Grades 2 through 5 (sixth grade was not served within the participating schools). Third, in contrast to Fuchs et al. (1997a), Saenz et al. sampled high-achieving (HA) classmates from each participating classroom so that 11 children were pre- and posttested from each class (all children participated): 2 LD, 3 LA, 3 AA, and 3 HA. PALS was implemented in English for 15 weeks, with strong fidelity observed and documented.

As with Fuchs et al. (1997a), results on the CRAB supported PALS efficacy; that is, on number of questions answered, PALS students outgrew control students, and the effect sizes were large: 1.06 for LD, 0.86 for LA, 0.60 for AA, and 1.02 for HA (across achievement status, 1.02). So, Saenz et al.'s findings extend those of prior work by broadening the generalization for PALS efficacy to include LEP students. In addition, this study extends findings to children who begin the program already reading better than their classmates, by showing that the reading comprehension of these initially high performers improves better in PALS classrooms that with conventional reading instruction.

Results Across Both Efficacy Studies

Across both efficacy studies, results demonstrate the potential of PALS to enhance children's reading comprehension performance. PALS's effectiveness may reside both in its specific activities as well as in its overall organization. In terms of substance, the PALS activities encourage students to practice research-based strategies, which have been shown to strengthen reading comprehension when implemented regularly on instructional level text. With respect to organization, PALS organizes highly structured, reciprocal, one-to-one interaction that provides all students with frequent opportunity to respond; facilitates immediate corrective feedback; increases academic engaged time; and offers social support and encouragement, with all students sharing the esteem associated with the tutoring role. Moreover, with the PALS scorecard system, students work cooperatively with partners but compete in teams to earn points. This keeps students working in a focused, productive, and constructive manner.

PALS also offers the advantage of implementation ease. PALS materials are concrete, specific, and user friendly—important criteria if practices are to be implemented (see McLaughlin, 1991, cited in Gersten, Vaughn, Deshler, & Schiller, 1995). A comprehensive teacher manual guides implementation. In addition, there is no need for teachers to develop additional materials or devote more time to reading instruction. Moreover, PALS can complement any instructional approach, including whole language instruction as well as explicit phonics. This is so because PALS enhances, rather than provides a radical substitute for, teachers' ongoing reading practices.

FIRST-GRADE PALS

Over the last 10 years, we have extended PALS downward to address the development of reading and reading-related skills at preschool, kindergarten, and first grade (see D. Fuchs & Fuchs, 2005, for a summary). At first grade, the structure of PALS parallels PALS at the higher grades, but the activities differ from those used at the higher grades.

Overview of First-Grade PALS

With First-Grade PALS, each class is divided into pairs based on rapid letter-naming performance. A higher and lower achieving student constitutes each pair. The higher performing student is always the Coach (tutor) first; when the pair completes an activity, the students switch roles and repeat the activity. Partners change every 4 weeks. In contrast to PALS at the higher grades, first-grade PALS begins with the teacher conducting 5 minutes of instruction, introducing new sounds and sight words, and then leading a brief segmenting and blending activity. Then students begin partner work, which comprises Sounds & Words activities and Partner Reading.

The first Sounds & Words activity is a letter–sound correspondence task, which lasts 3 minutes. The Coach points to a letter and prompts the Reader to say its sound. If the Reader makes a mistake or does not know the sound of a letter, the Coach uses a correction procedure. When the Reader has said all of the sounds, the Coach marks a happy face on the lesson sheet and 5 points on a point sheet. Then partners switch roles and repeat the activity.

The second Sounds & Words activity involves blending the 8 to 10 words used during teacher-directed instruction. The Coach prompts the Reader to sound out a word, and then the Coach tells the Reader to "Say it fast," and the Reader responds by reading the word. If the Reader makes a mistake, the Coach uses a correction procedure. When all 8 to 10 words have been blended, the Coach marks a happy face on the lesson sheet and 5 points on the point sheet. Then partners switch roles and repeat the activity. Sounding out lasts 5 minutes.

The third Sounds & Words activity is sight word practice. The Coach points to each word and prompts the Reader to read it by saying, "What word?" If the Reader says the wrong word, the Coach uses a correction procedure. The Coach marks a happy face and 5 points at the end of the activity. Then the partners switch roles and repeat the task. Sight word practice is conducted for 4 minutes.

In the fourth Sounds & Words activity, students practice reading decodable words and sight words in short stories. Before students read the story, the teacher introduces new "rocket" words and reviews old rocket words. Rocket words (e.g., *playground*, *birthday party*, and *office*) were added to the stories to increase interest value. Next, the teacher reads the story to the students, who follow along on their lesson sheets. The teacher emphasizes the importance of reading quickly and correctly. Coaches then prompt their Readers to read the story. Coaches use a correction procedure for errors. When the story has been completed, the Coach marks a happy face and 5 points, then partners switch roles and repeat the activity. The story activity lasts 5 minutes. Coaches and Readers mark a star on a chart if they have read the story the number of times the teacher designates (never to exceed three times). When all the stars on the chart are marked, the student receives a bookmark and a new chart.

After students have been doing Sounds & Words activities independently for 4 weeks, 10 minutes of Partner Reading is added. In Partner Reading, students apply decoding skills and sight word knowledge to connected text appropriate to their reading level. Teachers train students in Partner Reading procedures in two 20-minute sessions. The Coach reads the title of the book, pointing to each word. Then the Reader reads and points to the title. The Coach reads a page of the book and points to each word, then the Reader reads the same page, pointing to the words. Partners proceed through the book in this manner, then mark 5 points and repeat the process, switching roles.

Each book is read four times before partners trade it in for a new one. Partner Reading is conducted for 10 minutes.

During first-grade PALS, students mark and earn points for their team. The class is divided into two teams. Each pair on a team has a point sheet with boxes numbered from 1 to 200 on it. For every first-grade PALS activity they complete, the pair slashes a smiley face on their PALS lesson sheet, then slashes 5 points (boxes) on their point sheet. While monitoring students, the teacher awards additional points for following PALS rules, for good reading, and for using correction procedures. At the end of every week, each pair reports the number of points earned for the team. The teacher tallies the points for each team. The winning team is cheered, then the "losing" team is applauded for good effort. Teams change every 4 weeks.

In previous work (for a summary, see D. Fuchs & Fuchs, 2005), using study designs similar to those already described for PALS at the higher grades, first-grade PALS had been shown to effect superior decoding and word recognition among low performers with and without disabilities, average performers, and high performers, in high-poverty and middle-class schools. More recently (also see D. Fuchs & Fuchs, 2005), we were interested in extending first-grade PALS to address higher order skills. The goal was to determine whether the structure of PALS could be used to enhance first-grade fluency and comprehension.

Fluency Studies

As already discussed, repeated reading has been shown in others' work to promote fluency (e.g., Herman, 1985; O'Shea et al., 1987; Samuels, 1979). As typically studied in the research literature, however, repeated reading requires one-to-one adult supervision, making the intervention difficult to implement in regular classrooms, especially at first grade. Perhaps for this reason, little attention has been focused on the use of repeated reading as a developmental practice. Rather, its use has been limited to remediation, after students manifest reading fluency difficulties. We were interested in exploring whether the first-grade PALS structure could accommodate the use of repeated reading classwide, as a developmental rather than remedial practice to promote reading fluency for the full range of learners.

Toward that end, we designed a speed-reading game, which involved speeded repeated readings the sight word activity (i.e., the fourth activity within Sounds & Words) and the short story reading activity within first-grade PALS Sounds & Words. During the first 12 weeks of PALS, the speed-reading game was applied to the sight word activity within Sounds & Words; for the remaining 10 weeks, the speed-reading game was applied to the short story reading activity within Sounds & Words.

In the speed-reading game, the teacher reminds students about the importance of reading quickly and correctly, times the students as they read, and reminds students to return to the same starting point with each repeated reading. First, Coaches read (30-second readings when applied to sight words, 60-second readings when applied to short stories), starting at the beginning of the relevant Sounds & Words section each time. Readers listen to their Coaches and use a correction procedure, as needed. At the end of each timed segment, the Coach marks his or her initial next to the last word read. The goal is for the Coach to increase this number of words read during the second or third timed segment. After three timed readings, the students switch roles, and the Reader reads three timed segments while the Coach listens and corrects. If a student succeeds in reading more words in the second or third trial, he or she marks a star on a chart. When all the stars on the chart are marked, the student receives a bookmark and a new chart.

We conducted a series of two studies at first grade (Fuchs & Fuchs, 2005) to assess the added value of this classwide repeated-reading PALS procedure. The methodological structure of the studies was analogous to the studies already summarized in this chapter. In each study, we recruited 30 first-grade teachers from four Title I and four non-Title 1 schools. Within each school, we randomly assigned classrooms to three conditions: control (no PALS), PALS, and PALS with repeated reading (RR). In each classroom, for pre- and posttesting, we identified 16 students (8 low performers, 4 average performers, and 4 high performers on rapid letter naming, with teacher confirmation). A total of 385 children completed posttesting, 37 of whom had a disability. Teachers conducted PALS (with and without RR) for 20 weeks; fidelity of implementation was strong. Measures included letter sound fluency, segmentation, blending, Woodcock Reading Mastery Tests—Revised Word Attack and Word Identification, Weschler Individual Achievement Test—Spelling, and the Comprehensive Reading Assessment Battery words read correctly in 3 minutes (on a near-transfer and far-transfer passage) and questions answered correctly.

The results were as follows. On most measures of phonological awareness, decoding, and word recognition, students in both first-grade PALS conditions (with and without RR) performed statistically significantly better than control students. By contrast, on fluency and comprehension measures, only first-grade PALS with RR outperformed the control group. These effects applied to children in high-poverty and middle-class schools. They also applied to students with disabilities as well as low-, average-, and high-achieving classmates. The effect sizes on the fluency and comprehension outcomes approximated one third of a standard deviation. Although the effect size is small, it is reliable, with comparable effects demonstrated across separate studies conducted during 2 consecutive years. On the basis of this series of studies, we conclude that first-grade PALS is a feasible method for incorporating repeated reading into a classwide structure. This differs from the absence of reliable effects for RR over sustained reading at the higher grades (Simmons et al., 1994), suggesting the possibility that RR may be more relevant for the developmental needs of first-grade readers. In any case, we conclude that PALS with RR can be used productively with first-grade children to enhance fluency and comprehension outcomes, outcomes rarely targeted explicitly with such young children.

What About Targeting Strategic Reading Comprehension Skill Directly?

In case readers should think that the PALS research program has been uniformly successful, we finish this overview of PALS reading research program with a failed attempt to target strategic reading comprehension skill directly among first-graders (D. Fuchs & Fuchs, 2005). Research has documented the contribution of comprehension strategy instruction for students in the intermediate grades (e.g., Rosenshine & Meister, 1994; Slavin, 1994) as well as second and third grades (e.g., Fuchs et al., 1997; Saenz et al., 2005). Research on the efficacy of comprehension strategy instruction has not occurred at first grade, and specific methods to promote the development of reading comprehension strategies had not been developed.

To address this void, we designed peer-mediated comprehension strategy instruction for first-grade PALS. The first phase of this comprehension strategy instruction involved teacher-directed listening activities. The first several sessions involved listening comprehension of videos; thereafter, the listening comprehension activities concerned stories that the teacher read aloud. We began with listening comprehension because it permitted the use of relatively rich narratives that incorporate character, setting, plot, and other story features. These in turn facilitated more interesting and challenging comprehension activities that would have been impossible if the children were required to read and, as demonstrated by Kendeou et al. (see chap. 2, this volume), comprehension skills do transfer across modalities. Over time, the teacher gradually modified the way in which discussion about the meaning of the story occurred. First, she asked questions (described below) of individual students in a conventional fashion; then she directed questions to pairs of students; next, she had students ask questions of each other; and finally, pairs of children read to each other and asked questions of each other. After several months of listening comprehension, which eventually had students reading and asking questions of each other, the teacher transitioned the listening comprehension activity to reading.

Across the listening and reading phases, the comprehension strategy activity remained constant. It comprised three parts corresponding to before, during, and after listening. The first part involved activating prior knowledge and making predictions. The teacher (and eventually the PALS Coach) pointed to the book cover and illustrations and asked questions of the children about the picture. The purpose of these questions was to activate children's thinking about the book's focus and to encourage them to listen with purpose.

The second part was called "Think Time." The purpose of Think Time was to teach children the importance of thinking about what they're looking at (video); listening to (teacher reading); and, eventually, reading for themselves. Furthermore, Think Time was designed to promote the notion that thinking should be online all the time, dynamically connected to new information. Children were taught that comprehension is not something we do only after we see a video or read a book. To help children understand the idea of thinking interactively with story text, we used the metaphor of making a movie. As children listened/read, their job was to make a movie in their heads corresponding to what they were learning in the story. At least twice during story listening/reading, the teacher stopped the children and said "Think Time." The teacher then asked four questions: (a) "Who's the movie/story about now?"; (b) "Where are they now?"; (c) "What's happening in the movie/story now?"; and (d) "What might happen next?"

The third part involved expressing the gist—or "story shrinking," where the Reader's task was to summarize the story. The summarization had to include at least two story features, such as character and setting, or character and plot, or setting and outcome. The Coach's job was to monitor the Reader's summary for conciseness and completeness. If the Reader's summary included only one story feature, then the Coach prompted the Reader to include at least one other story dimension. The Coach was given a card that serves both as a prompt (with symbols representing the story features) and a checklist to help monitor the Reader's performance.

To study the effects of this first-grade PALS comprehension strategy instruction, we (D. Fuchs & Fuchs, 2005) conducted a randomized controlled trial in which 30 first-grade teachers were randomly assigned, within schools, to three conditions: (a) control, (b) PALS, or (c) PALS with comprehension strategy instruction. The design and measures paralleled the design and measures of the fluency studies just described. As documented consistently across the research program, first-grade PALS (with and without the comprehension strategy component) resulted in differentially superior improvement on word-level reading skills, including word attack, word identification, and spelling. It was interesting, however, that first-grade PALS plus comprehension strategy instruction was associated with less strong word-level skill than first-grade

PALS without the comprehension strategy instruction. Moreover, to our surprise, there was no added value for the comprehension strategy instruction on our outcome measure of comprehension (i.e., answering open-ended questions about information of high thematic importance from 400-word folk tales written at a first-grade readability level).

We speculated that the comprehension strategy practice "interrupted" the students' word- and text-reading practice, which may be more critical to the developmental reading stage of first-graders than are comprehension strategies. This is in keeping with van den Broek, Tzeng, Risden, Trabasso, and Basche (2001), who found that questioning during reading may interfere with whatever comprehension-struggling or beginning readers can muster. Of course, study findings served to prove the null hypothesis. As Kendeou et al. suggested (chap. 2, this volume), it is possible that our comprehension strategy was insufficiently based on the causal structure of the text (also see chap. 1, this volume), so the possibility remains that comprehension strategy instruction at first grade, using better designed interventions and/or superior measures of comprehension, may prove beneficial.

CONCLUSION

More than 10 years of randomized controlled studies, with real teachers implementing treatments with naturally constituted classrooms in actual schools, have demonstrated the potential for a simple peer-mediated structure such as PALS to enhance students' reading outcomes in important ways. These outcomes include fluency and comprehension at Grades 2 through 6; word-level skill, fluency, and comprehension at first grade; word-level skill at kindergarten; and prereading phonological skill and knowledge about the alphabet at preschool. On the basis of the PALS research program, we draw five conclusions.

First and most fundamentally, we conclude that at Grades 2 through 6, PALS is a promising platform for promoting strategic reading comprehension behavior. That is, using the PALS structure and organization, we have transformed research-based instructional procedures, developed in laboratorylike settings, into feasible practices that can be implemented by real teachers in real schools. Second, young children can be taught to work in pairs, using structured activities, so that they follow procedures accurately, with productive outcomes. Our third conclusion addresses the robustness of the PALS treatments. In the series of studies just described, no interaction between student type (disabled, low, average, high) and treatment condition was identified. Instead, where effects emerged, they pertained across student types. Moreover, those effects applied almost uniformly across Title I and non-Title 1 schools, and they pertained comparably across children who receive reading instruction in their primary language or not. This suggests the potential applicability of PALS across a wide range of settings for enhancing strategic reading comprehension behavior.

At the same time (and this brings us to our fourth conclusion), despite the robustness of findings, each and every study we have described in this chapter included students who failed to respond to the best of our treatments. Although space limitations preclude discussion of the prevalence and characteristics associated with unresponsiveness, our research program has systematically investigated these issues, with results reported elsewhere (e.g., D. Fuchs, Fuchs, & Compton, 2004; Fuchs et al., 1977a). Given this persistent finding in our work (and the work of others), it is important for practitioners to note that no validated practice is universally effective. For this reason, good practice demands ongoing monitoring of student progress so that nonresponders can be identified promptly and helped appropriately.

Our fifth and final conclusion concerns the need for caution. Our foray into first-grade reading comprehension was not successful. Students who received instruction limited to word-reading skill outperformed those whose PALS reading time was continually interrupted with activities designed to promote strategic reading comprehension skill. This suggests that first-grade reading comprehension strategy instruction may be unproductive when those children have underdeveloped word-level skills (also see van den Broek et al., 2001). Such a conclusion, of course, awaits additional studies that attempt intervention in alternative ways, perhaps relying more on the causal structure of text and/or individualizing the comprehension strategy to account for patterns of cognitive strengths and weaknesses (chap. 2, this volume; McNamara & Kintsch, 1996). In fact, we (D. Fuchs et al., 2005) are at present revising the first-grade comprehension strategy intervention and will soon test the efficacy of that revised approach with the hope of promoting first-grade development on strategic reading comprehension skill.

ACKNOWLEDGMENTS

Support for the research described in this chapter was provided by Grants G008730253, H023E90020, H023C40001, and H324V980001 from the Office of Special Education Programs in the U.S. Department of Education and Core, and by Grant HD 15052 from the National Institute of Child Health and Development to Vanderbilt University. The content in this chapter does not reflect the positions or policies of these funding agencies, and no official endorsement by them should be inferred.

REFERENCES

Allen, V. (1976). *Children as teachers: Theory and research on tutoring*. New York: Academic.

Baker, J. M., & Zigmond, N. (1990). Are regular education classes equipped to accommodate students with learning disabilities? *Exceptional Children*, *56*, 515–526.

Baumann, J. F. (1984). The effectiveness of a direct instruction paradigm for teaching main idea comprehension. *Reading Research Quarterly*, *16*, 32–35.

Bean, T.W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Reading Behavior*, *16*, 297–307.

Bierman, K. L., & Furman, W. (1981). Effects of role and assignment rationale on attitudes formed during peer tutoring. *Journal of Educational Psychology*, *73*, 33–40.

Braddock, J., II, Hawley, W., Hunt, T., Oakes, J., Slavin, R., & Wheelock, A. (1992). Realizing our nation's diversity as an opportunity: Alternatives to sorting America's children. Nashville, TN: Vanderbilt Institute for Public Policy Studies, Center for Education and Human Development Policy.

Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior*, *22*, 1–14.

Brown, A. L., & Smiley, S. (1977). Rating the importance of structural units of prose passages: A problem of metacognitive development. *Child Development*, *48*, 1–8.

Cohen, E. (1984). Talking and working together: Status, interaction, and learning. In P. Peterson, L. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction* (pp. 171–186). New York: Academic.

Collins, J., & Smith, E. E. (1982). Teaching the process of reading comprehension. In D. K. Detterman & R. J. Sternberg (Eds.), *How and how much can intelligence be increased* (pp. 173–185). Norwood, NJ: Ablex.

Doctorow, M., Wittrock, M. C., & Marks, C. (1978). Generative processes in reading comprehension. *Journal of Educational Psychology*, *70*, 109–118.

Englert, C. S., & Hiebert, E. H. (1984). Children's developing awareness of text structures in expository materials. *Journal of Educational Psychology*, *76*, 65–75.

Fuchs, D., & Fuchs, L. S. (1994). Inclusive schools movement and the radicalization of special education reform. *Exceptional Children*, *60*, 294–309.

Fuchs, D., & Fuchs, L. S. (2005). Peer-assisted learning strategies: Promoting word recognition, fluency, and reading comprehension in young children. *Journal of Special Education*, *39*, 34–44.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2004). Identifying reading disabilities by responsiveness-to-instruction: Specifying measures and criteria. *Learning Disability Quarterly*, *27*, 216–227.

Fuchs, D., Fuchs, L. S., & Compton, D. (2005). *Response to intervention for determining risk, preventing, and identifying reading disability*. Grant proposal. Available from first author.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1996). *Peer-assisted learning strategies in reading: A manual*. Available from Box 328 Peabody, Vanderbilt University, Nashville, TN 37203.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997a). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, *34*, 174–206.

Fuchs, D., Fuchs, L. S., & Thompson, A. (1997b). *Evaluation of first-grade PALS Title I implementation in Nashville*. Unpublished manuscript.

Fuchs, L. S., & Fuchs, D. (2003). Can diagnostic reading assessment enhance general educators' instructional differentiation and student learning? In B. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 325–351). Timonium, MD: York.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989). Monitoring reading growth using student recalls: Effects of two teacher feedback systems. *Journal of Educational Research*, *83*, 103–111.

Fuchs, L. S., Fuchs, D., Phillips, N. B., & Simmons, D. (1993). Contextual variables affecting instructional adaptation for difficult-to-teach students. *Exceptional Children*, *22*, 725–740.

Gersten, R., Vaughn, S., Deshler, D., & Schiller, E. (1995). *What we know (and still don't know) about utilizing research findings to improve practice: Implications for special education.* Unpublished manuscript.

Grasser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.

Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology*, *81*, 371–383.

Hare, V. C., & Borchardt, K. M. (1984). Direct instruction of summarization skills. *Reading Research Quarterly*, *20*, 62–78.

Herman, P. A. (1985). The effect of repeated reading on reading rate, speech pauses, and word recognition. *Reading Research Quarterly*, *22*, 553–565.

Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, *56*, 473–493.

Hodgkinson, H. L. (1995). What should we call people?: Race, class, and the census for 2000. *Phi Delta Kappan*, *77*, 173–179.

Jenkins, J. R., Heliotis, J. D., Haynes, M. C., & Beck, K. (1986). Does passive learning account for disabled readers' comprehension deficits in ordinal reading situations? *Learning Disabilities Quarterly*, *9*, 69–75.

Markman, E. M. (1981). Comprehension monitoring. In W. P. Dickson (Ed.), *Children's oral communication skills* (pp. 61–84). New York: Academic.

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–287.

O'Shea, L. J., Sindelar, P. T., & O'Shea, D. J. (1987). The effects of repeated reading and attentional cues on the reading fluency and comprehension of learning disabled readers. *Learning Disabilities Research*, *2*, 103–109.

Palincsar, A. M., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *2*, 117–175.

Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology*, *76*, 1239–1252.

Posner, M. I., & Snyder, C. R. R. (1975a). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55–85). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Posner, M. I., & Snyder, C. R. R. (1975b). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbit & S. Dornic (Eds.), *Attention and performance* (Vol. 5, pp. 669–682). New York: Academic.

Pressley, M. (1997). Remarks on reading comprehension. Notes prepared for the Chesapeake Institute, Washington, DC.

Rinehart, S. D., Stahl, S. A., & Erickson, L. G. (1986). Some effects of summarization training on reading and studying. *Reading Research Quarterly*, *21*, 422–438.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of research. *Review of Educational Research*, *64*, 479–530.

Saenz, L., Fuchs, L. S., & Fuchs, D. (2005). Effects of peer-assisted learning strategies on English language learners: A randomized controlled study. *Exceptional Children*, *71*, 235–247.

Samuels, S. J. (1979). The method of repeated readings. *Reading Teacher*, *32*, 403–408.

Simmons, D., Fuchs, D., Fuchs, L. S., Pate, J., & Mathes, P. (1994). Importance of instructional complexity and role reciprocity to classwide peer tutoring. *Learning Disabilities Research and Practice*, *9*, 203–212.

Slavin, R. E. (1994). *Cooperative learning: Theory, research, & practice* (2nd ed.). Boston: Allyn & Bacon.

Stevens, R. J., Madden, N. A., Slavin, R. E., & Farnish, A. M. (1987). Cooperative integrated reading and composition: Two field experiments. *Reading Research Quarterly*, *22*, 433–454.

Top, B. I. L., & Osguthorpe, R. T. (1987). Reverse-role tutoring: The effects of handicapped students tutoring regular class students. *Elementary School Journal*, *87*, 413–423.

van den Broek, P., Tzeng, Y., Risden, K., Trabasso, T., & Basche, P. (2001). Inferential questioning: Effects of narrative texts as a function of grade and timing. *Journal of Educational Psychology*, *93*, 521–529.

Wiegmann, D. A., Dansereau, D. F., & Patterson, M. E. (1992). Cooperative learning: Effects of role playing and ability on performance. *Journal of Experimental Education*, *60*, 109–116.

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*, 345–376.

8

Literacy in the Curriculum: Integrating Text Structure and Content Area Instruction

Joanna P. Williams
Teachers College, Columbia University

Reading in the elementary school curriculum involves expository text, for the most part, but this type of text is often neglected in the early grades. This chapter briefly reviews the research, most of it dealing with students at or above the fourth grade, which demonstrates the importance of text structure in reading comprehension and the effectiveness of text structure instruction. This literature provides the basis for the development of an instructional program designed for younger children, who, it has been shown, can benefit from explicit instruction in comprehension. The program, which focuses intensively on one specific expository structure, compare/contrast, is described. Then the results of evaluations of the program are presented, followed by a discussion of further work designed to refine and expand it.

Many children do not easily understand what they read, and there are many reasons for such difficulties. Sometimes there is lack of fluency, and cognitive resources that might be allocated to comprehension must be given over to basic word recognition processes. The evidence for this statement comes from the high correlations that have been found between oral reading fluency measures and standardized measures of reading comprehension (Lyon, Fletcher, Fuchs, & Chhabra, in press). Another factor that contributes to comprehension difficulties is a lack of vocabulary knowledge and general background knowledge, which children acquire both from reading and from their general interaction with the world during their early years (Bos & Anders, 1990).

A third common difficulty arises because successful reading involves persistence. Torgesen (1977) noted that some students are inactive as learners. Even when they are taught specific ways in which they could get more out of their reading (e.g., how to underline), these students do not greatly improve their reading performance. Fourth, when reading is difficult, students experience frustration, which lowers their motivation. They begin to avoid reading and so lose opportunities for further practice and achievement (Stanovich, 1986).

In response to these observations and analyses, remedial and preventive programs have been developed in decoding, vocabulary, general knowledge (often through listening activities), and cognitive strategies. This last includes instruction in cognitive monitoring and specific procedures to use when one's monitoring indicates that one is not comprehending adequately.

This chapter focuses on another factor that leads to comprehension difficulties: the lack of sensitivity to text structure. Most textual information is content information. Readers use that information to construct a meaningful mental representation of the text. However, some textual information concerns structure, not content. This structural information is important because it helps readers organize the content and thus aids in the process of constructing a mental representation, that is, the meaning of the text.

Text structure is inherent in a text's organizational pattern, which reflects the logical connections among the ideas in the text (Meyer, Brandt, & Bluth, 1980; Meyer & Poon, 2001; chap. 14, this volume). These patterns are not limited to text; they represent general rhetorical structures (Dickson, Simmons, & Kameenui, 1998). There may be explicit markers in the surface text that guide the reader, such as signal words and phrases like *first* or "as a result." These words identify the particular genre of the text (narrative or expository) and the particular type of structure within a genre. There may also be titles or headings that cue the overall organization of the text. Proficient readers have a sense of the structures that exist, so that even when there are no surface cues to the text's structure and even with text that is not organized effectively, they can organize the information presented in text into a well-structured mental representation.

Discussions of text structure instruction are often subsumed under the category of cognitive strategy instruction (see chap. 1, this volume). The rationale for teaching comprehension strategies is that readers derive more meaning from text when they engage in intentional thinking. That is, when people run into difficulties understanding what they have read, the application of specific strategic cognitive processes will improve their comprehension. Studies have shown that instruction in comprehension strategies is effective in helping students learn strategies and that when strategies are applied, better comprehension follows (Pressley & McCormick, 1995). The goal for this instruction is for the reader to internalize the strategy so that its use becomes automatic. In situations in which comprehension is a problem, the strategy can be brought to consciousness and applied.

With respect to text structure, the goal is to make students aware that text has structure, to familiarize them with the cues that exist in text, and to provide sufficient practice so that they can respond to those cues when they are reading. We also want students to be able to apply their knowledge of structure to reading situations in which the text is not well organized, that is, to reorganize the text and give it the structure that will make it easier to understand. Ideally this will happen automatically; however, if it does not, then the students should be able to apply their knowledge consciously, knowing that they are using a cognitive strategy.

IMPORTANCE OF TEXT STRUCTURE

There is an extensive literature that demonstrates that well-structured text enhances recall and comprehension (Pearson & Dole, 1987), and many studies have shown that instruction designed to teach students to recognize the underlying structure of text improves comprehension (Gersten, Fuchs, Williams, & Baker, 2001). In such instruction, students are taught to identify the important structural elements of a particular type of text and then to memorize a list of generic questions that cue a search for those important elements. The students thus acquire knowledge about text and begin to use this strategically.

Different types of texts are organized in different ways. Narrative text typically follows a single general structural pattern (often called *story grammar*; Mandler & Johnson, 1977). Children develop sensitivity to narrative structure early, and they use it to comprehend simple stories before they enter school. That is, they note the setting, the main character, the important conflicts (actions and reactions of the characters) and the story resolution as they read. However, this structure encompasses narrative only at the plot level, and many stories have meaning beyond the plot level. Mature comprehension involves generalization beyond the story characters and events to real life people and events. This requires comprehension at the level of the story theme (Williams, Brown, Silverstein, & deCani, 1994; Williams et al., 2002; Wilder & Williams, 2001).

Expository text, on the other hand, comes in a variety of patterns. For example, Meyer (Meyer et al., 1980; chap. 14, this volume) has listed six: (e.g., description, sequence, listing, compare/contrast, cause–effect, and problem solution). Because of this fact, and because it often deals with unfamiliar content, expository text is generally more difficult to comprehend (Kucan & Beck, 1997).

People sometimes argue that because only a small proportion of authentic text actually follows any single specific structure, there is little reason to spend much instructional time on text structure. It is true that for some children the reading experience attained over the first years of schooling will be sufficient for them to attain sensitivity to structure. For many children, however, this is not enough. I suggest that early comprehension instruction is likely to be more effective if it includes specific instruction in text structure, including the use of texts that are well structured and prepared specifically for particular instructional purposes.

Of course, these specific structures are not limited to text; they are rhetorical structures that reflect universal cognitive processes. The thinking of young children exhibits forms of all these structures. By the time children enter school, they tell stories, compare and contrast objects, order events in a temporal sequence, and attribute causality (Carey, 1990). But children have not had sufficient experience to be able to use these structures with ease, and sometimes they do not even recognize opportunities for using them to enhance their comprehension. The expectation is that helping students to recognize the structure inherent in text—and match it to their own cognitive structures—will help them understand and produce not only text but also spoken discourse. Then, when they encounter text whose structure is complex or text that is poorly organized, they will be able to simplify or reorganize it in order to better comprehend it.

INSTRUCTIONAL EFFORTS

There is a rather substantial literature concerning instruction about expository text structure (see chap. 14, this volume), but there are not many studies that deal with young children. There is almost no instruction focused on expository text in the early grades. In fact, until recently there has been little exposure to expository text in these grades. Hoffman et al. (1994) pointed out that basal readers typically include a very small amount of expository text. Duke (2000), who examined first-grade classrooms across several school districts, found a scarcity of informational texts in all of them. It is likely that this lack of experience with expository text contributes to the fourth-grade slump in reading achievement noted by Chall, Jacobs, and Baldwin (1990).

In an instructional study conducted by Armbuster, Anderson, and Ostertag (1987), middle school students were given either explicit instruction in the problem solution structure or more traditional instruction that included general comprehension questions and summarization. The students who received explicit instruction recalled more information on an essay test than did others. The structure-trained students also identified more main ideas than did the other students, indicating that explicit instruction in structure facilitates the development of a well-structured mental representation.

Richgels, McGee, Lomax, and Sheard (1987) found that sixth-graders were indeed sensitive to and aware of text structure and that their awareness varied as a function of structure type. Across five awareness and recall tasks, students were more consistently aware of a compare/contrast structure than a causation structure. Thus, upper elementary school students are promising candidates for instruction in text structure.

Overall, the few instructional studies that exist, although far from conclusive, suggest that instruction, especially if geared to a single text structure, is effective in improving students' ability to comprehend expository text. Dickson (1999), for example, found that the compare/contrast structure could be taught successfully in middle school general education classrooms. Much more work in this area needs to be done. Very few instructional programs have been developed, and there is almost no work that focuses on the sustainability of effects over time, or on transfer effects. Without proper attention to expository text in the early grades, students remain unprepared for the comprehension demands that await them (Bernhardt, Destino, Kamil, & Rodriguez-Muñoz, 1995; Jitendra, Hoppes & Xin, 2000).

Much of the work that I have cited here was done by researchers in special education who were investigating the problems of children with learning disabilities. More recently, researchers in general education have found that many of the difficulties exhibited by children with learning disabilities are also seen in poor readers who do not have learning disabilities. It appears that most of the instructional techniques first explored with students with learning disabilities can also be of great help with developing instruction for other children at risk for academic failure.

THE CASE FOR EXPLICIT INSTRUCTION IN COMPREHENSION

The theoretical base in the area of reading comprehension has been developed only recently (RAND Reading Study Group, 2002), and over the past decades comprehension instruction in the schools has been quite unstructured (Williams, 1998). It often incorporates strategic instruction of a highly metacognitive nature, emphasizing reflection and self-monitoring. This instruction has been effective at the middle school level and above (Allington, Guice, Michelson, Baker, & Li, 1996). However, such an approach is contraindicated for younger children and those at risk for academic failure. I argue that a structured and explicit approach is required for these children. In such an approach, the overall conceptualization of a reading strategy remains the same, but it is interpreted more directly; that is, more emphasis is given to text signals and patterns and less to metacognitive processing.

This same structured and explicit approach has been found to be effective in teaching other academic content to at-risk children. It follows classic principles of good instructional design, that is, content is introduced in small increments, moving from the simple to the complex, providing modeling by the teacher; scaffolding that fades as instruction progresses; and at each step, substantial practice with feedback, first guided and then independent. In the same spirit, instructional materials (texts) are simple; in fact, they are sometimes developed specifically for the instruction, so that they exemplify with clarity the particular textual patterns that are the focus of instruction.

There remains the question of when it is most appropriate to start such instruction. Children must have an adequate level of basic cognitive abilities, including reasoning and memory, in order to benefit from comprehension instruction. They must also have some facility with word recognition and decoding, in order to deal with reading comprehension. Some investigators, including Palincsar and Brown (1984), have suggested that because many of the same cognitive abilities are the basis for understanding spoken language, it would make sense to introduce children to comprehension instruction before they can actually read texts independently, by having the teacher read stories aloud and conducting group discussions.

Chapters 2 and 3 of this volume, both of which focus on narrative text, present data that indicate that reading comprehension ability is predicted by early language comprehension. In chapter 3, Oakhill and Cain show specifically that understanding story structure at an early age is an important predictor of later comprehension.) Both of these research teams suggest providing targeted comprehension instruction before beginning reading instruction.

Pappas (1993) and Duke and Kays (1998) have shown that emerging readers recognize expository language and can recall the content of expository trade books, and they recommended that young children be provided with more opportunities to listen to expository text in the primary grades. The National Reading Council (Snow, Burns, & Griffin, 1998) recommended that there be more exposure to expository text in the early years of schooling, and Dreher and Baker (2005) explored the effects of classrooms' use of informational books in Grades 2 to 4.

Smolkin and Donovan (2001, 2003) have used Krashen's (1976) monitor theory of second-language acquisition and learning to conceptualize reading comprehension across the elementary grades. Krashen suggested that there are two knowledge systems that underlie children's second language performance. The first is *acquisition*, which operates in an unconscious fashion. The second, *learning*, operates during formal instruction. Smolkin and Donovan consider that the stage of emergent literacy represents Krashen's acquisition system; children pick up some awareness of print simply through exposure to their environment. In first grade, they move beyond this to learning letter–sound relationships, writing, and so forth, which they would not typically pick up themselves without being instructed. These tasks are based on simple associative learning, of which first graders are developmentally capable, and so the formal instruction is effective.

With respect to comprehension, Smolkin and Donovan (2003) argued that at the first grade level, and into the second grade, children are not developmentally ready for formal learning. However, they are ready for comprehension acquisition, that is, they have responsiveness to texts read to them, and they can benefit from informal classroom read-alouds and discussions.

My colleagues and I believe that the greater attention recently given to expository text in the elementary grades is a good step forward, but it may be that the informal read-alouds and discussion that are the core of most primary classrooms are insufficient. Moreover, it is not clear that systematic comprehension instruction is not effective at that level; it has not been tried. My colleagues and I believe that the same classic principles of instruction that have been found to be so effective in word recognition (as well as other basic skills, e.g., elementary arithmetic; Fuchs et al., 2004) can be useful in teaching comprehension in the primary grades and that the results of the studies presented in this chapter indicate the value of explicit, structured instruction at the second-grade level (Williams, 2003; Williams, Hall, & Lauer, 2004).

OUR PROGRAM: CLOSE ANALYSIS OF TEXT WITH STRUCTURE

Study 1: Development and Evaluation of the Program

My colleagues and I have developed an instructional program for second-graders whose goal is to improve the reading comprehension of compare/contrast expository text (Williams et al., 2005). We taught students three strategies: (a) clue words to identify a text as a compare/contrast text, (b) a graphic organizer to lay out the relevant information in the text, and (c) a series of questions that would help them focus on the important information in the text. We also emphasized the close analysis of specially written well-structured exemplars of the compare/contrast text structure.

We chose animal classification as the program content; our goal was to teach students the characteristic features of each of the five classes of vertebrates (mammals, birds, fish, reptiles, and amphibians). We included one animal as a prototypical example of each of the five classes (lion, eagle, shark, crocodile, frog). This content is included in the standards for elementary-level science curricula in New York State.

We used a trade book about each of the five animals and a comprehensive animal encyclopedia. We also prepared short target paragraphs to be read and analyzed. Each of these paragraphs contained several comparative statements about two of the five animals; these statements presented the information that was the basis for categorizing them into five vertebrate classes. These paragraphs became longer as the program proceeded. Toward the end of the program, they also included some distractor information, that is, general information about one or the other of the two animals that could not be put together with any other information in the paragraph to construct a comparative statement.

Here are two examples of target paragraphs:

(1) Eagles and crocodiles are wild animals. Eagles are warm-blooded; however, crocodiles are cold-blooded. Eagles and crocodiles both lay eggs.

(2) Frogs and crocodiles are interesting animals. Frogs and crocodiles are alike; they are cold-blooded. Frogs can jump very far and they have long sticky tongues. Frogs have smooth skin, but crocodiles have scales. Crocodiles use their sharp teeth to protect themselves. Crocodiles get oxygen to breathe from the air. They can't breathe underwater. However, frogs get oxygen to breathe from the air and the water.

The program consisted of nine lessons, which were taught in 15 sessions. Each lesson focused on two of the five prototypical animals and contained the following seven sections: (a) clue words, (b) trade book reading and discussion, (c) vocabulary development, (d) reading and analysis of target paragraph, (e) graphic organizer, (f) compare/contrast strategy questions, (g) summary (with a paragraph frame as a support); and (h) lesson review. The first lesson focused on two very familiar animals (cats and dogs) to help introduce students to the procedure without being distracted by new content.

Clue Words. At the beginning of each lesson, the teacher previewed the purpose of the lesson and introduced eight words (*alike*, *both*, *and*, *compare*, *but*, *however*, *than*, and *contrast*). The teacher wrote the clue words on the board and elicited sentences that used one of the clue words.

Trade Book Reading and Discussion. During the next part of the lesson, teachers read to the class about the two animals from the encyclopedia and the trade books. Teachers then directed a discussion about the animals. This part of the lesson provided information about the animals beyond the specific information contained in the target paragraphs. It was also designed to heighten motivation—which is particularly important because difficulty in comprehending expository text may, in part, be attributed to lack of student interest (Armbuster et al., 1987).

Vocabulary Development. Teachers then introduced vocabulary concepts related to animal classification (oxygen, *hair*, *scales*, *feathers*, *warm blooded*, *cold blooded*).

Reading and Analysis of Target Paragraph. The students read the target paragraph silently, and then the teacher reread it as students followed along on their own copy. Students then analyzed the text to narrow in on the similarities and differences found in the paragraph. Students identified the individual sentences that represented the similarities and differences. They then circled all the clue words. Finally, they took turns generating sentences that described how the two animals in the paragraph were the same or different. The teacher encouraged them to use well-structured comparative statements, that is, sentences that were based on accurate information from the paragraph that included a clue word.

Graphic Organizer. Next, students organized the paragraph's content with the help of a matrix, the graphic organizer that best represents compare/contrast structure (Chambliss & Calfee, 1998). An individual matrix was used for each animal feature that was compared in the paragraph. Students then wrote a well-structured comparative statement to match the content organized in the matrix. Paragraphs in earlier lessons contained less information (and therefore there were fewer matrices) than paragraphs in later lessons.

Compare/Contrast Questions. The students then organized the statements they had generated according to the following three questions: (a) What two things is this paragraph about?, (b) How are they the same?, and (c) How are they different?

Summary. Next, students wrote summaries of the paragraph. Summarization skills are complex, so students were provided with just a paragraph frame to use as a prompt. This structured approach to writing is particularly helpful to young children who are just beginning to develop their writing skills (Harris & Graham, 1999). In the later lessons, no frame was provided.

Review. At the end of each lesson the teacher and students reviewed the vocabulary and the strategies (clue words, graphic organizer, and compare/contrast questions).

An overview of the Text Structure program, along with an overview of the comparison content program, which is described below, is presented in Table 8.1.

Evaluation of the Text Structure Program. Reading comprehension instruction is typically focused on content, not structure. My colleagues and I

TABLE 8.1
Program Overview

Text Structure Program	*Content program*
Clue words	Background knowledge
Trade book reading and discussion	Trade book reading and discussion
Vocabulary development	Information web
Reading and analysis of target paragraph	Vocabulary development
Graphic organizer (matrix)	Reading of target paragraph
Compare/contrast questions	General content discussion
Summary (paragraph frame)	Summary (paragraph frame)
Lesson review	Lesson review

wanted to compare our program to one having the same content so that we could evaluate the specific effects of adding instruction concerning structure. Therefore, we developed a second program that taught the same content, using the same instructional materials (trade books, encyclopedia, and target paragraphs), but that did not emphasize text structure. As a control, we also looked at students who received neither program. The main purpose of the study was to determine whether instruction focused on text structure would help second-grade students improve their comprehension of compare/contrast expository text. We also had a further question. The school day contains a finite amount of time, and choices must be made as to how to use that time. If teaching students about text structure means that they will learn less content, then we must be prepared to make a trade-off. A better outcome would be that there was no decrease in content learning. Therefore we asked whether this type of instruction on text structure would detract from students' ability to learn new content.

Teachers of 10 second-grade classes in three New York City public schools volunteered to participate. We randomly assigned the 10 intact classes to one of the three treatments (Text Structure: $n = 4$; content: $n = 4$, and no instruction: $n = 2$). A total of 128 students participated. Across the three schools, the enrollment included approximately 56% Hispanic children, 41% African American, 2% Caucasian, and 1% Asian. Almost 90% of the children received state aid in the form of free or reduced-rate lunch. Approximately 6% of the students were enrolled in special education services.

The Content Program. The comparison content program was designed to correspond to more traditional content-area instruction and was intended to be a viable program; students participating in this program would, we believed, learn important content that would enable them to comprehend novel paragraphs about similar content.

As in the Text Structure program, there were 15 sessions, so that overall the same amount of time was given to the instruction. Each lesson consisted of the following eight sections: (a) background knowledge, (b) trade book reading and discussion, (c) information web (a graphic organizer that organizes information topically), (d) vocabulary development, (e) a reading of target paragraph, (f) general content discussion, (g) summary (with paragraph frame), and (h) lesson review.

Results. A summary of the posttest findings, which are described here, is presented in Table 8.2.

Strategy Measures. After the lessons, we interviewed students individually, asking them to respond to questions both orally and in writing. First, we included several measures that evaluated the acquisition of the three strategies taught in the Text Structure program. We assessed recall of clue words; ability to locate the clue words in a paragraph; ability to generate sentences (oral and written) based on information they had graphically organized; and, finally, recall of the three compare/contrast questions. On the first four of these measures, the students who received the Text Structure instruction did significantly better than the students in the other two groups. On the fifth measure, recall of the three compare/contrast questions, there was no effect of treatment.

The comparison content program included one strategy: a graphic organizer, that is, an information web. There were no differences among the three treatment groups in their proficiency in this strategy; all groups achieved relatively high scores. This finding was not surprising. By the second grade, most students are acquainted with the web strategy.

Outcome Measures. We looked at two types of outcome measures. The first type addressed our comprehension goals; these measures assessed students' ability to gain information from expository text. The second type evaluated the content goals, examining what specific information about animals the students had learned.

First, we looked at *comprehension*. We examined the students' ability to summarize a compare/contrast paragraph that contained material explicitly taught in the program, that is, information about two of the five instructed animals. The test paragraph compared two animals that had been directly compared during the instruction. We asked for written summaries. We counted the number of summary statements that were both accurate and included an appropriate clue word. The Text Structure group performed better than did the other two groups, although the differences were not significant.

Then we investigated the students' ability to *transfer*. The goal of reading comprehension instruction is to have students improve in their ability to read novel content, not simply to reread material they have already practiced.

TABLE 8.2
Posttest Performance by Classroom Condition, Study 1

Outcome measure	Test Structure (TS; n = 4)		Content (C; n = 4)		No instruction (N; n = 2)		Effect size[a]	
	M	SD	M	SD	M	SD		
Strategy measures								
							TS > C	TS > N
Recall of clue words (maximum = 8)	4.33	1.63	0.62	0.49	0.28	0.01	3.50**	4.94**
Locating clue words (maximum = 6)	4.61	1.15	0.42	0.33	0.47	0.15	5.66***	6.37***
Matrix sentence generation: Oral (maximum = 2)	1.28	0.52	0.59	0.26	0.60	0.03	1.77*	2.43*
Matrix sentence generation: Written (maximum = 3)	1.28	0.46	0.69	0.12	0.71	0.15	2.03*	1.84*
Recall of compare-contrast questions (maximum = 3)	0.51	0.25	0.34	0.20	0.20	0.02		
Information Web (maximum = 3)	2.27	0.15	2.13	0.43	2.33	0.34		
Comprehension measures								
							TS > C	TS > N
Explicit teaching: Written (instructed animals: maximum = 3)	0.82	0.51	0.34	0.09	0.27	0.13	1.60§	1.72§
Immediate transfer: Oral (instructed) animals/novel combo; maximum = 6	2.50	0.51	1.56	0.31	1.39	0.35	2.29*	2.58*
Near transfer: Oral (novel animals; maximum = 6)	2.14	0.72	1.13	0.35	0.81	0.02	1.91*	3.60*
Far transfer: Oral (unrelated content; maximum = 6)	1.83	0.59	1.04	0.24	1.11	0.15	1.93*	1.95*
Structure transfer: Oral (pro/con; maximum = 3)	1.41	0.38	1.21	0.34	1.56	0.37		
Content measures								
							TS > N	C > N
Vocabulary concepts (maximum = 6)	3.58	0.70	3.46	1.02	2.13	0.59	2.23***	1.66*
Detail Questions (maximum = 5)	2.05	0.36	2.70	0.51	1.71	0.29		

[a]Effect sizes are shown for only significant effects.
§$p < .10$, *$p < .05$; **$p < .01$, ***$p < .001$.

Therefore, we developed a series of three compare/contrast texts that were structured in the same way as those used in the instruction. However, the content was different. In each of these three paragraphs, the content became further removed from the content used in the instruction.

First, there was a paragraph that contained information about two of the instructed animals, but two that had not been directly compared in any of the lesson target paragraphs (sharks and crocodiles). We considered this a measure of what was explicitly taught in the program. The second paragraph also contained information concerning animal classification, but it contained information about two animals that had not been mentioned during the instruction (elephants and turtles). We considered this a measure of *near transfer*. The last paragraph in the series contained information unrelated to animal classification; it compared bikes and cars. We considered this a measure of *far transfer*. Children provided oral summaries.

We found that, across the three paragraphs, the Text Structure groups scored significantly higher than either of the two groups not only on the explicitly taught paragraph but also on the transfer paragraphs. These findings indicated that the Text Structure students had in fact transferred what they had learned. It is not particularly surprising to find that after reading comprehension instruction, students do better on tests that involve the same material on which they were instructed. However, it is not very common to find positive effects of the instruction when the tests involve new material not seen during instruction.

The second type of outcome measure focused on how much of the content about animals had been learned. We were interested in two types of content learning: (a) vocabulary and (b) factual details. First, the data indicated that students learned the vocabulary concepts (e.g., *oxygen* and *warm blooded*) that they were taught. On this measure, we found a different pattern of results from what we found on the text structure outcome measures. The Text Structure group attained a higher score on the vocabulary measure than did the content group; in addition, the content group (who also received the vocabulary concepts instruction) did better than the no-treatment control group. In our second content outcome measure, we asked the students one detail question about each of the five target animals. These questions tapped information that was presented in the read-alouds that were the basis for the class discussions. Here, there was no effect of instructional group.

These findings concerning the amount of content learning are important because they indicate that spending substantial instructional time on text structure training did not detract from the amount of content the students learned. The content group, whose instruction focused solely on content, did not acquire more information about animals than the Text Structure group did.

Our final outcome measure looked at transfer of another type, the ability to *transfer to a new text structure* (called *structure transfer* in Table 8.2). We asked whether, if second-graders were given highly structured, very intensive training in one text structure, they would also read and comprehend another type of expository text better than would noninstructed students.

We did not have a hypothesis that there would be transfer to a second structure. There are no prior data on this point. We chose pro/con as the second structure, because both compare/contrast and pro/con conform to a simple associative structure in which information is categorized under either of two headings (same and different or for and against). We provided nothing in our instruction that would serve as a bridge to the pro/con posttest task, as would be recommended for instruction designed to maximize transfer (Salomon & Perkins, 1989), especially for young students (Fuchs et al., 2003). However, given that our study was an initial step in developing and evaluating this type of explicit comprehension instruction, we thought that it would be useful to explore the potential of our program for structure transfer as well as for content transfer.

The paragraph followed a pro/con structure, including arguments for and against keeping animals in zoos. As in the other transfer tasks, the students read the paragraph and provided an oral summary. There were no differences among the three instructional groups. These results indicate that instruction in one particular text structure does not automatically transfer to other text structures, even if the two structures are related.

Individual Differences in Responsiveness to Instruction

We looked further, beyond comparison of classroom means, to determine how well individual children were responding to the instruction. On the basis of certain of the posttest outcome measures that assessed strategy application and transfer, we categorized the participants as either responders or nonresponders (Al Otaiba & Fuchs, 2002). We established two sets of criteria for acceptable performance on these measures, one stringent and the other more lenient. Regardless of which criteria we used, we found that the students who had not performed as well as the others had lower listening and lower reading comprehension scores. However, there was no relationship between nonresponding and special education status; that is, it was no more likely that a student with an individualized education plan or a referral for one would be nonresponder than a student who was in neither of these categories.

Study 2: Refining and Extending the Text Structure Program

The first goal of our second study was to refine the program and replicate our findings (Williams, Stafford, Lauer, Hall, & Pollini, 2006). We made some small modifications in the program in response to teachers' comments or to our own classroom observations. We simplified the graphic organizer, which had turned out to be cumbersome to teach, and we put more emphasis on the three strategy questions.

We made another, more important, modification. In our first study, when posttest items required a summary of what had just been read, we provided the students with a paragraph frame to help guide their summaries. This corresponded to the design of the final lessons in the program, which used such a frame. In the revised version of the program, we took the instruction further: We removed this scaffold at the midpoint of the program. After that point, students had to generate their summaries without the support of the frame. Our posttest paragraphs also did not include a frame; they required a more independent response than had been required in our initial study.

We also extended our work in two additional ways. In our first study, all test paragraphs were written to conform to the clear compare/contrast pattern that we were attempting to teach. But a large proportion of real life text is not well structured, and that is one of the reasons typically given for not teaching text structure in the classroom. Thus, we decided that it was important to establish a transfer effect to authentic text, so in our second study we added to the posttest a compare/contrast paragraph that was taken from an authentic trade book written at the second-grade level and that focused on the instructional content (animals). Here is the paragraph:

> Frogs are usually found around marshes, ponds, or other wet places. Toads live mostly on dry land. Frogs have shiny smooth skin. Toads have rough, bumpy skin. Both are often brownish or dark green in color, with stripes or marking in other colors. But some frogs are bright green, blue, white, yellow, or red. Frogs can jump farther than toads. Many frogs have webbed feet that help them swim. Tree frogs have special pads on their toes that help them climb. Most toads and frogs lay their eggs in water.

We reflected on our finding that although our initial evaluation of the program had shown transfer to different content, it had not shown any transfer to the related pro/con structure. So, we asked whether the introduction of a small amount of explicit training on this second structure would lead to improvement in performance. If so, that would suggest that we might be able to develop a text structure program that would encompass more than one expository structure without involving a substantial amount of additional training. We introduced four pro/con clue words—*good*, *advantage*, *bad*, *disadvantage*—into the final three lessons (six class sessions) of the program. This was the only strategy that was included for pro/con comprehension. The target paragraphs on which the students did their text analysis during these three lessons were of a mixed structure; that is, they contained both compare/contrast and pro/con sentences. Because of this mixture, they more clearly approximated authentic text.

The revised version of the program consisted of 12 lessons taught in twenty-two 45-minute sessions. Fifteen second-grade classrooms (the total number of students was 215) in schools similar to those in the first study were randomly assigned to the Text Structure, content, and no treatment groups. The results were clear. This time, the Text Structure group was significantly better on all three of the text structure strategies (clue words, graphic organizer, and strategy questions). Again, all three groups performed at a high level, and not differently, on the content strategy (the information web).

The results on all the outcome measures concerning our comprehension goals were replicated. The Text Structure students were better than the other students on constructing oral summaries, including summaries of novel texts in transfer tasks. Moreover, this time they were also significantly better on their written summaries. In addition, we found that there was transfer to authentic text, although the level of performance was not as high as it was in the other test paragraphs that conformed more closely to the structure of the paragraphs used in instruction. Also, the students who received the Text Structure program and those who received the content program were not significantly different on either of the content outcome measures.

We found that adding a limited amount of instruction on the pro/con structure was effective; performance on the pro/con test item was better for Text Structure students. Scores were not as high, however, as scores on the compare/contrast text paragraphs. This finding indicates that there could be some savings in an instructional program that covered a variety of expository text structures. However, these results suggest that an instructional program would have to devote a substantial amount of practice to each structure—or at least to the second one taught—if high performance levels are to be achieved.

Study 3: Incorporating Content Goals Into the Text Structure Program

We believe that our first two development/evaluation cycles offer convincing evidence that we have an effective way to teach text structure, especially given the fact that our Text Structure program does not detract from the amount of content acquired. However, up to this point we had not addressed the issue of just how much content could be acquired in the context of a program that was primarily focused on instruction about structure. The purpose of the next study (Williams, 2006) was to revise the program to ensure that students would not only achieve our text structure goals but also learn a specified amount of content, that is, information about animal classification.

We narrowed our focus to encompass both our structure and new content goals within the time permitted for classroom evaluation (about 10 weeks of instruction). We decreased the number of classes of vertebrate from five to three (mammal, bird, and reptile). This allowed us to develop lessons that incorporated substantial focus on the content as well as on text structure. We presented six animals, two from each class (lions, bears, turtles, crocodiles, eagles, and parrots) over the course of the program.

Our focus on close reading and analysis of target compare/contrast paragraphs remained. We added three classification paragraphs that described the features that make an animal a mammal, a bird, or a reptile; the ones that were relevant to each lesson's target animals were reviewed in that lesson. This gave the students practice on the features that define each vertebrate class.

In addition, we changed the graphic organizer from an activity designed to teach a strategy for text structure analysis to one that supported our new content goals. This decision to modify the graphic organizer was based on teachers' comments and on our classroom observations. We felt that the other two structure strategies (the clue words and the strategy questions) were sufficient to lead students to produce well-structured and accurate comparative statements in their summaries. We did, however, wish to include a graphic organizer in the program; children respond well to such activities. So we included one that was designed to reinforce the content of the program. Also, to focus on our main research question, we eliminated the instruction on the pro/con structure.

We randomly assigned classrooms to the Text Structure program (n = 4 classrooms), the content program (n = 4), and to the no-instruction program (n = 3). A total of 173 students participated. The students were drawn from schools whose demographics were similar to those of the earlier studies, except that there were more Hispanic students (51%) than African Americans (46%) in this study. The programs consisted of nine lessons taught in nineteen 45-minute sessions.

We found that our revised Text Structure program, even with its heavy dose of content instruction, significantly improved the comprehension of compare/contrast text. We respect to strategy acquisition, performance on the text structure strategy measures indicated very high scores on the part of the Text Structure group. As in the earlier study, there was no significant difference among the groups on the content graphic organizer; all students did very well.

With respect to the outcome measures, the Text Structure condition was significantly superior to the other two conditions on both the oral and the written comprehension measures (ability to provide well-structured and accurate comparative sentences in their summaries, including transfer measures containing content both related and unrelated to the instructional text).

We also found positive effects with respect to the content goals. We added a third content measure, which we called *classification*: We assessed recall of the specific information about the animal features that had been taught in the program. On both vocabulary concepts and classification measures, both the Text Structure and the content groups performed significantly better than did the no-treatment group; there was no difference between the Text Structure and the content-only groups. Thus, the Text Structure program did not detract from the students' acquisition of the content that represented the focus of the content goals. However, the results on the third content measure, detail questions, showed a different pattern of performance. This measure, which tapped content tangential to the program's content goals, showed significantly superior performance on the part of the Content students.

These findings indicate that instruction in text structure that also seeks to impart a serious amount of content does not reduce the acquisition of the content that is central to the content goals of the program. A program that focuses only on content acquisition, however, may lead to acquisition of more tangential content along with the central content of the program. Of course, there will be no improvement in knowledge of text structure from such a program.

CONCLUSIONS

Instructional designs must have as their goal the development of a range of methods and techniques so that all children will have an opportunity to achieve. It is our opinion, however, that most, if not all, methods and techniques that will be found to be successful will involve explicit, structured instruction.

Our next goal is to use what we have learned to develop instruction in another important content area: social studies. We have replicated our positive findings with that type of content (Williams, Nubla-Kung, Pollini, Stafford, Garcia, & Snyder, in press). We also hope to pursue the question of how much intensive instruction must be given in each of the several basic expository structures; it may be that sufficient instruction on a variety of structures is the key to achieving good performance on authentic (ill-structured) text. There are also other questions that we wish to address. We do not know how long lasting the effects of our instruction would be. It has not been feasible for us to evaluate the sustainability, because in New York City the pressure arising from the need to administer standardized tests in the third grade has kept us from getting permission to give delayed posttests.

Although my colleagues and I have concentrated on cognitive outcomes, we acknowledge the importance of motivation and engagement issues. One of our goals in developing our instruction was to provide interesting content and a variety of appealing instructional activities. In all of our studies we have observed the affective responses of the students. Typically, the children's attention is well focused during the lessons, and we rate them as highly engaged. Motivation and engagement are important for the teachers as well. Debriefing interviews indicate that teachers like our Text Structure program, especially its explicitness and the built-in repetition and reviews. They also like the way we have organized the instruction and feel that the program is easy to administer. They concur with our observations that their students respond positively to the instruction. We have been gratified by the number of teachers who have elected to continue to use the program after their involvement in our research is over.

We support the recommendation of the National Reading Council (Snow et al., 1998) that there be a greater presence of expository text in primary-grade classrooms. We go further, though: We believe that these children would benefit from explicit instruction in expository text structure; it would help prepare them for the reading challenges that they will encounter in the later grades.

REFERENCES

Allington, R., Guice, S., Michelson, N., Baker, K., & Li, S. (1996). Literature-based curricula in high poverty schools. In M. F. Graves, P. van den Broek, & B. M. Taylor (Eds.), *The first R: Every child's right to read*. New York: Teachers College Press.

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention. *Remedial and Special Education*, *23*, 300–316.

Armbuster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly*, *22*, 331–346.

Bernhardt, E., Destino, T., Kamil, M., & Rodriguez-Muñoz, M. (1995). Assessing science knowledge in an English/Spanish bilingual elementary school. *Cognosos*, *4*, 4–6.

Bos, C. S., & Anders, P. L. (1990). Effects of interactive vocabulary instruction on the vocabulary learning and reading comprehension of junior-high learning disabled students. *Learning Disability Quarterly*, *12*, 31–42.

Carey, S. (1990). Cognitive development. In D. N. Osherson & E. E. Smith (Eds.), *Thinking: An invitation to cognitive science* (Vol. 3, pp. 147–172). Cambridge, MA: MIT Press.

Chall, J. S., Jacobs, V. A., & Baldwin, L. E. (1990). *The reading crisis: Why poor children fall behind*. Cambridge, MA: Harvard University Press.

Chambliss, M. J., & Calfee, R. C. (1998). *Textbooks for learning*. Malden, MA: Blackwell.

Dickson, S. (1999). Integrating reading and writing to teach compare–contrast text structure: A research-based methodology. *Reading & Writing Quarterly*, *14*, 49–79.

Dickson, S. V., Simmons, D. C., & Kameenui, E. J. (1998). Text organization: Research bases. In D. C. Simmons & E. J. Kameenui (Eds.), *What reading research tells us about children with diverse learning needs* (pp. 239–277). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Dreher, M. J., & Baker, L. (2005, June). *Balancing learning to read and reading for learning: Infusing information books into primary-grade classrooms*. Paper presented at the meeting of the Society for the Scientific Study of Reading, Toronto, Ontario, Canada.

Duke, N. K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, *35*, 202–224.

Duke, N. K., & Kays, J. (1998). "Can I say 'once upon a time?'" Kindergarten children developing knowledge of information book language. *Early Language Research Quarterly*, *13*, 295–318.

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., & Schroeder, K. (2003). Enhancing third-grade students' mathematical problem solving with self-regulated learning strategies. *Journal of Educational Psychology*, *95*, 306–315.

Fuchs, L. S., Fuchs, D., Prentice, K., Hamlett, C. L., Finelli, R., & Courey, S. J. (2004). Enhancing mathematical problem-solving among third-grade students with schema-based instruction. *Journal of Educational Psychology*, *96*, 635–647.

Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, *71*, 279–230.

Harris, K. R., & Graham, S. (1999). Programmatic intervention research: Illustrations from the evolution of self-regulated strategy development. *Learning Disability Quarterly*, *22*, 251–262.

Hoffman, J. V., McCarthy, S. J., Abbott, J., Christian, C., Corman, L., Curry, C., et al. (1994). So what's new in the new basals? A focus on first grade. *Journal of Reading Behavior*, *26*, 47–73.

Jitendra, A. K., Hoppes, M. K., & Xin, Y. P. (2000). Enhancing main idea comprehension for students with learning problems: The role of summarization strategy and self-monitoring instruction. *Journal of Special Education*, *34*, 127–139.

Krashen, S. (1976). Formal and informal linguistic environments in language acquisition and language learning. *TESOL Quarterly*, *10*, 157–168.

Kucan, L., & Beck, I. L., (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, *67*, 271–299.

Lyon, G. R., Fletcher, J., Fuchs, L. & Chhabra, V. (in press). Treatment of learning disabilities. In E. Mash & R. Barkley (Eds.), Treatment of childhood disorders (3rd ed.). New York: Guilford.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111–151.

Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly*, *16*, 72–103.

Meyer, B. J. F, & Poon, L. W. (2001). Effects of the structure strategy and signaling on recall of the text. *Journal of Educational Psychology*, *93*, 141–159.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*, 117–175.

Pappas, C. C. (1993). Is narrative "primary"? Some insights from kindergarteners' pretend reading of stories and information books. *Journal of Reading Behavior*, 25, 97–129.

Pearson, P. D., & Dole, J. A. (1987). Explicit comprehension instruction: A review of research and a new conceptualization of instruction. *Elementary School Journal*, *88*, 151–165.

Pressley, M., & McCormick, C. (1995). *Advanced educational psychology*. New York: Harcourt Brace.

RAND Reading Study Group. (2002). *Reading for understanding*. Santa Monica, CA: RAND.

Richgels, D. J., McGee, L. M., Lomax, R. G., & Sheard, C. (1987). Awareness of four text structures: Effects on recall of expository text. *Reading Research Quarterly*, *22*, 177–196.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist*, *24*, 113–142.

Smolkin, L. B., & Donovan, C. A. (2001). The contexts of comprehension: The information book read aloud, comprehension acquisition and comprehension instruction in a first-grade classroom. *Elementary School Journal*, *102*, 97–129.

Smolkin, L. B., & Donovan, C. A. (2003). Supporting comprehension acquisition for emerging and struggling readers: The interactive information book read-aloud. *Exceptionality*, *11*, 25–38.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Stanovich, K. E. (1986). Cognitive processes and the reading problems of learning-disabled children: Evaluating the assumption of specificity. In J. K. Torgesen & B. Y. L. Wong (Eds.), *Psychological and educational perspectives on LD* (pp. 87–131). Orlando, FL: Academic.

Torgesen, J. (1977). The role of nonspecific factors in the test performance of learning disabled children: A theoretical assessment. *Journal of Learning Disabilities*, *10*, 33–39.

Wilder, A. A., & Williams, J. P. (2001). Students with severe learning disabilities can learn higher order comprehension skills. *Journal of Educational Psychology*, *93*, 268–278.

Williams, J. P. (1998). Improving the comprehension of disabled readers. *Annals of Dyslexia*, *48*, 213–238.

Williams, J. P. (2003). Teaching text structure to improve reading comprehension. In H. L. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 295–305). New York: Guilford.

Williams, J. P. (2006). *Teaching reading comprehension in the context of elementary science instruction*. Manuscript in preparation.

Williams, J. P., Brown, L. G., Silverstein, A. K., & de Cani, J. S. (1994). An instructional program for adolescents with learning disabilities in the comprehension of narrative themes. *Learning Disabilities Quarterly*, *17*, 205–221.

Williams, J. P., Hall, K. M., & Lauer, K. D. (2004). Building the basics of comprehension instruction: Teaching expository text to young at-risk learners. *Exceptionality*, *12*, 129–144.

Williams, J. P., Hall, K. M., Lauer, K. D., Stafford, K. B., DeSisto, L. A., & de Cani, J. S. (2005). Expository text comprehension in the primary grade classroom. *Journal of Educational Psychology*, *97*, 538–550.

Williams, J. P., Lauer, K. D., Hall, K. M., Lord, K. M., Gugga, S. S., Bak, S. J., et al. (2002). Teaching elementary school students to identify story themes. *Journal of Educational Psychology*, *94*, 235–248.

Williams, J. P., Nubla-Kung, A. M., Pollini, S., Stafford, K. B., Garcia, A., & Snyder, A. E. (in press). Teaching cause/effect text structure through social studies content to at-risk second graders. *Journal of Learning Disabilities*.

Williams, J. P., Stafford, K. B., Lauer, K. D., Hall, K. M., & Pollini, S. (2006). *Embedding reading comprehension training into content-area instruction*. Manuscript submitted for publication.

9

What Brains Are For: Action, Meaning, and Reading Comprehension

Arthur M. Glenberg, Beth Jaworski, and Michal Rischal
University of Wisconsin—Madison

Joel Levin
University of Arizona

The authors argue that brains evolved to control action, and, as suggested by M. Montessori (1967), a successful theory of cognition and its application will require recognition of that fact. The indexical hypothesis, an embodied account of language comprehension, posits that language is understood by simulating the actions that underlie sentence meaning and that reading comprehension can be improved by ensuring that this simulation occurs. Accordingly, first- and second-grade children manipulated toys (e.g., animals on a toy farm) to correspond to sentences in the stories that they were reading. Compared with children in a reread control group, manipulation improved memory and comprehension by almost 2 standard deviations. Similar improvements were found when the children were instructed to imagine the manipulation. In another experiment, manipulation and imagined manipulation produced a large increase in the ability of 3rd-grade children to solve some (but not all) mathematical story problems.

What are brains for? The answer to that question would seem to be obvious: Brains are for thinking! But consider some of the exquisite work that the brain does that one would not ordinarily classify as thinking. The brain controls and coordinates eye and neck muscles that allow us to track an annoying housefly (not to mention the tremendous amount of work done by the visual areas of the brain in locating and identifying the fly). The brain coordinates that tracking activity with reaching out with the hand and arm to swat the fly. Finally, the brain simultaneously controls and coordinates virtually every muscle in the body so that the path we walk will intersect the fly's trajectory while reaching with the arm and tracking with the eyes. It is a good bet that the brain coordinates control of movement much more frequently (and successfully) than thinking. Thus, the answer to the question "What are brains for?" might well be: Brains are for action!

The noted biologist Rudolfo Llinas puts it this way: "The nervous system is only necessary for multi-cellular creatures … that can orchestrate and express active movement" (Llinas, 2001, p 15). A particularly fascinating illustration of this fact is the life cycle of the sea squirt, a tunicate member of the phylum chordata (animals with spinal cords, including humans). The sea squirt begins life as an egg and then has a short existence (a few hours to a few days) as a tadpole-like creature with a primitive spinal cord, a brain, a light-sensitive organ, and the ability to express active movement. The sea squirt moves about until it finds a suitable attachment site, such as a rock. After attaching, the sea squirt never leaves; that is, it never again expresses active movement. The fascinating part of the story is that almost as soon as the sea squirt stops moving, it begins to ingest its brain! No action, no need for a brain.

Is the sea squirt simply an oddity, or does it have implications for human brains and human cognition? Of course, brains partake of many functions in addition to action, including thinking without expressing overt movement. Nonetheless, the notion that brains are for action provides an organizing theme, or inspiration, for conceptualizing cognitive functions that have been divorced from action in many contemporary theories. Thus, the plan for this chapter is to take the "brains are for action" theme all the way from the sea squirt to human cognition and then into the classroom. We begin with speculation about action and the evolution of cognition, and then we discuss the intimate connection between action and language comprehension. That connection leads to the development of a strategy to improve reading comprehension both when reading in the service of text comprehension and when reading to accomplish other tasks, such as mathematical story problem solving. Finally, we will discuss the practicality of transferring these ideas to the classroom.

ACTION, EVOLUTION OF COGNITION, AND MEANING

Imagine a chipmunk emerging from a hole in the ground to survey the environment. Upon seeing a threat, such as a snake, an effective action for the chipmunk is to dive back into its hole. If the chipmunk attempted to fly away instead, it would be consumed by the snake and not make much of a contribution to the gene pool. In contrast, suppose that it were a robin that sees the snake. Effective action for the robin is to fly away, whereas if it attempted to dive into a hole, it would be dead. Finally, effective action for a human who cannot fly or dive into a hole may be to flick away the snake with a long stick. Although cognition has surely advanced from the bird to the human, in all cases when the cognitive system is determining effective action it must take into account the body morphology of the actor. Thus, it is likely that there has been co-evolution of cognition and the body; that is, the evolution of one is constrained by the evolution of the other.

In addition to contributing to effective action, we also suppose that cognition is what makes meaning, or sense, out of the world. Is there any way in which action and meaning can be connected? Glenberg (1997) proposed that meaning is tied to action. Namely, the meaning of a situation to an individual (human or nonhuman animal) consists of the set of actions the individual can undertake in that situation. The set of actions is determined by the goal-directed mesh of affordances. The notion of an *affordance* was introduced by the perceptual psychologist James Gibson (1979). The term captures an interaction between body morphology and the physical environment. That is, what can be done with a physical object, such as a chair, depends on the body of the animal interacting with it. (See Glenberg, 1997, for an extension of the Gibsonian meaning of affordance to cover nonprojectable features, that is, learned aspects of interaction. For example, the fact that a chair is a museum piece may not change the Gibsonian affordance that one may sit in it; however, learning has endowed this situation with additional information, retrieved from memory, that prevents sitting.) For an adult human, a kitchen chair affords sitting, standing on, or even lifting to flick a snake off a porch with one of the chair's legs. For a toddler, the chair affords sitting and standing on, as well as hiding under in a game of hide and seek. The chair does not afford lifting for the toddler, however, because the toddler is not strong enough to lift it. Because the affordances of the chair differ for the adult and the child, the claim is that the meaning of the chair also differs. *Mesh* refers to the process by which affordances are combined to accomplish goals. The mesh process must respect both body and physics. Thus, an adult can mesh the affordances of a kitchen chair (e.g., it can be lifted to expose a leg) and the affordances of a porch (e.g., it affords secure footing) to accomplish the goal of flicking a snake off the porch. It is hard to imagine, however, how the adult could mesh the affordances of a beanbag chair to accomplish the same goal; hence beanbag chairs have a different meaning.

This action-based account of meaning seems to be far from our everyday sense of meaning, that is, the sort of meaning conveyed by language. In fact, however, the data (to be reviewed shortly) point to a strong connection, and the *indexical hypothesis* (IH) was developed by Glenberg and Robertson (1999, 2000) to provide a theoretical description of that connection. The IH postulates that three, temporally overlapping processes—(a) indexing, (b) derivation of affordances, and (c) meshing as directed by grammar—are used to convert words and sentences into embodied, action-based meaning. That is, we go from the words to a consideration of the actions implied by the sentence; if we can create a smooth and coherent (i.e., doable) simulation, then we can understand the sentence.

Consider how a person might understand a sentence such as "Art flicked the snake out of the way using the chair." The indexing process maps words (e.g., "Art") and phrases (e.g., "the chair") to objects in the environment or to perceptual symbols (Barsalou, 1999). Perceptual symbols are neurally based representations of objects and events that are extracted from perceptual stimulation by selective attention. Thus, on seeing a kitchen chair being used as step stool, one might attend to its height, the fact that it has a flat seat, and that the back is situated to afford grasping and balancing while standing on the seat. These aspects of the representation maintain their neural basis in that they are stored in areas of cortex associated with the initial perceptions. Thus, the information regarding the flat seat is stored in visual areas, whereas the ability to grasp the back is stored in motor areas (see discussion by Hauk, Johnsrude, & Pulvermüller, 2004). Later, if the chair has just been varnished, the smell of the chair will be stored in olfactory areas and linked together with the other aspects of the chair representation. With sufficient experience, these representations take on the properties of simulators (Barsalou, 1999); that is, the analogically encoded properties of a chair can be cognitively combined with other representations, such as that of standing or hefting or hiding. So, according to the first process of the IH, on reading the sentence, one probably indexes "Art" to the perceptual symbol of a male human and "the chair" to the perceptual symbol of, perhaps, a kitchen chair.

According to the IH, affordances are derived from the objects (or their perceptual symbols), not the words. Note that words do not have affordances in the usual sense; that is, people do not interact with words as they do physical objects. Instead, it is the objects that words designate that have affordances and are a major source of meaning As noted by Barsalou (1999) and Pulvermüller (in press), word forms (e.g., the sound of the word *chair* or the way the printed word looks) have perceptual characteristics stored in auditory and visual areas of the brain, but affordances come from associating those word forms with perceptual experiences of the referents. (See Glenberg & Robertson, 2000, and Kaschak & Glenberg, 2000, for demonstrations that understanding both familiar and unfamiliar concepts depends on accessing perceptual information, not just putative abstract concepts.)

The third, temporally overlapping process specified by the IH is that grammatical knowledge of several sorts is used to mesh or combine the affordances into a coherent simulation. For example, the syntax of "Art flicked the snake out of the way using the chair" indicates that Art did the flicking, not the snake. In addition, knowledge of the meaning of the syntactic construction sets the goal for the mesh process, that is, what the meshing must accomplish. For example, the first part of the sentence (up to *using*) is in the form of the caused motion verb-argument construction (Goldberg, 1995). Goldberg (1995) proposed that this syntactic construction (subject–verb–object–locational phrase) conveys the meaning that the subject ("Art") causes the motion of the object "the snake" to a location ("out of the way"). Kaschak and Glenberg (2000) demonstrated that English speakers are sensitive to these sorts of constructional meanings and that speakers use the meaning of the constructions to help determine the meaning of the individual words within the construction. Other sorts of grammatical knowledge (e.g., the meaning of temporal adverbs such as *while* and *after*) direct the temporal course of the meshing process (De Vega, Robertson, Glenberg, Kaschak, & Rinck, 2004).

The IH proposes that a sentence is understood to the extent that the meshing process results in a coherent, that is, doable, simulation of action. Consider what would happen, however, if you happen to index "chair" to your favorite beanbag chair. Beanbag chairs do not afford (easy) lifting, let alone the flicking away of snakes. In this case, you might consider the sentence to be nonsense or the speaker to be misinformed. The point is that the same words in the same sentence might be deemed perfectly sensible or nonsense depending on how those words are indexed, the affordances derived, and whether those affordances can be meshed as directed by syntax.

A tremendous amount of evidence is accumulating that supports the notion that linguistic meaning is based on the body's perception and action systems, that is, the systems that derive and mesh affordances. For example, Hauk et al. (2004) recorded brain activation while participants listened to verbs such as *pick*, *kick*, and *lick*. While listening to *pick*, the brain was especially active in that portion of the motor strip that controls finger movements, whereas when listening to *kick*, the activation was greatest in the portion of the motor strip controlling the leg. As another example, Borghi, Glenberg, and Kaschak (2004) had participants read sentences such as "There is a car in front of you" and then verify whether probe words (e.g., *roof*, *wheel*, or *road*) named a part of the object mentioned in the sentence (e.g., *car*). In one condition, participants responded "yes" by moving the arm upward to reach a response button labeled *yes*, and they moved downward to reach a button labeled *no*. In the other condition, people responded "yes" by moving to the lower button and "no" by moving to the upper button. What would be expected if the mere reading of a word such as *roof* activated action-based meaning? Because the roof of a car is near the car's top, interacting with a roof requires movement upward, and the mere reading of the word should prepare one to act in an upward manner. Consistent with this prediction, people responded faster to words such as *roof* when the yes response required an upward movement than when it required a downward movement. Just the opposite was found for words like *wheel*. A more complete review of these data can be found in Glenberg (in press).

ACTION AND ABSTRACTION

The IH might work for concrete objects and actions such as "wheel" and "kick"; however, any theory of language and conceptualization must also be able to account for what appear to be abstract ideas such as "p or not-p," "truth," and grammatical knowledge. Furthermore, such a theory must account for how differently abled people (e.g., people with disabilities) can understand. A complete discussion of these issues would be well beyond the mandate of this chapter, but it may be useful at least to sketch some possibilities. Lakoff (1987) discussed how logical predicates, such as "p or not-p," might be understood. He began by noting that some types of human activities, such as dealing with containers, are virtually universal. From activities such as filling containers, drinking from them, putting the hand inside and outside, and so on, we develop, in Lakoff's terminology, image schemas, which are similar to Baraslou's (1999) simulators. It is important to note that the activity of interacting with containers is structured by how the body works. For example, a finger can be inside the container, or outside the container, but rarely both. Thus, the understanding of a container becomes structured with the equivalent of "things can be inside or outside, but not both." According to Lakoff, logical rules such as "p or not-p" are understood by analogy to concrete experiences with the right structure, such as the container simulator. (For another example, see Barsalou's discussion of how concepts such as "truth" and "or" can be built from perceptual symbols.)

Glenberg and Kaschak (2002; see also Glenberg & Gallese, 2006) discussed how grammatical knowledge can be built from action. Consider our understanding that the grammatical form S-V-O-O implies transfer. Thus, "Art gave Danielle the book," "Art sent Danielle the book," and even "Art blicked Danielle the book" all seem to indicate that Danielle got the book from Art by different means, even if we are uncertain what those means may be. Tomasello (2000, 2003) has documented how children first acquiring their native language learn verb islands such as "give me X." These islands coalesce around frequent exemplars of verbs, such as *give*, and only with extensive experience do children develop the full adult S-V-O-O construction (Casenhiser & Goldberg, 2005). Now, imagine a baby who is first learning the meaning of *give* by having an adult say "Give me the cup." While making the utterance, the adult may literally stretch the baby's arm and open the baby's hand. The actions of stretching out and opening the hand then become the meaning of "give X." As the baby hears more language and can use that language to direct attention to several objects in the environment, the baby can learn to understand more complex forms, such as "Give Momma the cup"; that is, *give* initiates an arm action that will end with a hand opening to release whatever is in the hand (e.g., a cup). On hearing the word *Momma*, the baby's attention is directed toward the mother, and the affordances of the mother are perceived. Some of those affordances are related to the fact that Momma can also grasp the cup (that she is an appropriate recipient) and that she is in a particular location. We suppose that the child learns to use the location of Momma to guide the arm action corresponding to giving. Thus, the grammatical structure of the construction corresponds to the structure used to control action associated with transfer (Glenberg & Gallese, 2006).

The action-based understanding of the abstract concept of transfer is exactly what is implied by experimental data. Glenberg and Kaschak (2002) found that requiring an arm movement in opposition to arm extension interferes with understanding sentences such as "Give Danielle the book." Furthermore, such an arm movement also interferes with the understanding of sentences such as "Tell Danielle the secret"; that is, the abstract notion of transfer of information is grounded in the simple action of transfer of objects. In fact, this same sort of mechanism appears to be operating in the understanding of counterfactuals such as "If she had been more discreet, Art would have told Danielle the secret," as demonstrated by De Vega (in press).

Within an action-based account of language and cognition, there are good reasons to suspect that differently abled people will develop concepts similar to others; that is, a differently abled person will have perception and action systems substantially similar to the population norm and hence will interact with the world in substantially similar ways. Thus, a congenitally blind person will be able to understand "p or not-p" much as a sighted person because the blind person interacts with containers in a similar manner. Nonetheless, the account predicts specific, albeit subtle, deficits that are beginning to be found. For example, Neininger and Pulvermüller (2003) investigated language understanding in patients with lesions in areas controlling action or vision. The patients did not report any language difficulties, and there were few deficits on standard neuropsychological assessments (and no indication of aphasia). Nonetheless, with the speed stress of a lexical decision task, the patients with lesions in areas controlling action made more errors on action words than did patients with lesions in visual areas, and the reverse was found for errors on words referring to visual properties. Thus, differences in perceptual and action systems do influence language understanding, although for two reasons, apparently large differences in abilities (e.g., the blind person or the person with paraplegia) need not have correspondingly large differences in understanding. First, when one system functions differently, there will be substantial overlap among people in the great variety of other perceptual and action systems, thus constraining how situations and language can be understood. Second, patients often have ample opportunity after trauma to develop alternative routes for understanding (e.g., after damage to visual areas, the patient might use proprioception to recover spatial information that might otherwise be based on vision.)

ACTION AND LEARNING TO READ

Why are some children verbally skilled when it comes to conversation, but hate to read? For some children, the problem may be poor fluency due to inadequate practice (e.g., Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001), or failure to develop powerful reading strategies as described in many of the chapters in this volume. We suspect, however, that part of the problem is related to indexing. Consider the situation when a child is first exposed to language as a baby. Almost all of the words are immediately indexed to objects and actions (Masur, 1997). For example, when a mother says, "Here is your bottle," she is actually holding the bottle and directing the child's attention to the object and the actions associated with that object. Or, when a father says, "Wave bye-bye," he demonstrates how to wave. The point is that when learning an oral language, indexing is frequent and immediate, reinforcing the connection between words and the objects or their perceptual symbols. Consider now the situation when a child is learning to read. By necessity, the child must focus on the arduous process of decoding, that is, producing sounds from the letters. Thus, the child's attention is explicitly drawn away from the to-be-indexed objects to the letters of the word. Furthermore, the relevant objects are unlikely to be present in the environment, and there is certainly none of the gesture found in conversation. Even if a child is successful in pronouncing the word by combining the sounds of letters (e.g., "duh awh guh"), the blending of the sounds and their prosody is much different from the smooth production found in conversation (e.g., "dog"). Because of the differences in the sound of spoken words compared to

Breakfast on the farm

Ben needs to feed the animals.

He pushes the hay down the hole.

The goat eats the hay.

Ben gets eggs from the chicken.

He puts the eggs in the cart.

He gives the pumpkins to the pig.

All the animals are happy now.

Figure 9.1. Top: Example of a story from the farm scenario with "green light" signals. Bottom: The farm scenario toys.

sounded-out words, the latter is an unreliable cue to the perceptual symbol of the object. In short, we suspect that some children find reading to be an unrewarding activity because they have not learned to successfully index the written word. For those children, reading is an effortful and boring exercise in word calling.

Glenberg, Gutierrez, Levin, Japuntich, and Kaschak (2004) developed an early-reading intervention designed to ensure indexing of written words to objects. First- and second-grade children read short stories about scenarios,

TABLE 9.1
Effects of Physical and Imagined Manipulation

Type of manipulation	*Day 1 (story1) physical*	*Day 1 (story2) imagine*	*Day 2 imagine*	*Day 3 no instruction*
Group	**Proportion recall of action sentences**			
Manipulate	.62	.66	.63	.70
Re-read	.29	.31	.24	.47
	Proportion correct on spatial inference questions			
Manipulate	.93	.83	.89	.86
Re-read	.76	.72	.64	.75

such as stories about life on a farm. In front of the child was a set of toys (see Figure 9.1) to which the words could be indexed. In the *physical manipulate* condition, children read aloud critical sentences (those signaled by a picture of a green traffic light) and then were taught to move the toys to correspond to the sentences. This procedure virtually guarantees that the child indexes the words to objects and meshes the affordances in order to correctly move the toys. In the *reread* condition, children also read about the scenario, and the toys were in front of them. However, when these children saw a green light, they were to reread the sentence. After reading and a 2-minute interval filled with distracting conversation, the children were asked to recall as much of the story as possible, and they were asked a series of inference questions that required integration of text and scenario knowledge.

As shown in the first column of Table 9.1, the children who used physical manipulation recalled more, and more successfully answered the inference questions, than children who read and reread the critical sentences. The effect size (Cohen's *d*) for recall was 1.39, and for answering the inference question it was 0.81. In other words, the effects were substantial.

After applying physical manipulation, the children were taught to imagine manipulating the objects; that is, they were told to figure out how they would move the objects, but instead of actually moving them, they were to imagine moving them. In the reread condition, the children were taught to read the text once out loud and once silently. Columns 2 and 3 in Table 9.1 show that the benefits of manipulation extend to imagined manipulation. That is, children do not have to always physically manipulate; once they learn how to index, the indexing can be done in imagination, much the way we suppose that competent adults read. The effect sizes for imagined manipulation compared to reread were 1.87 and 1.50 for the recall and question answering, respectively. On the third day of the experiment, children were asked to read texts (that had the green lights), but they were not prompted to use any particular strategy. The finding (fourth column in Table 9.1) that children who had previously used physical manipulation and imagined manipulation were still comprehending and remembering more than children who had reread demonstrates that the imagined manipulation strategy is readily maintained (d = 1.23 for recall).

The results from two other experiments further demonstrate the efficacy of the manipulation technique. Glenberg, Brown, and Levin (in press) had children participate in groups of three. One child would read a sentence and manipulate the toys while the others watched, then the next child would read and manipulate, and so on. The results indicated that the benefits of manipulation compared to rereading extended to watching other children manipulate ($d = 1.72$). Marley, Levin, and Glenberg (2005) used the manipulation technique with Native American children with learning disabilities. The children listened to the story and then observed the outcome of the experimenter manipulating, or thought about each sentence as it was presented. Much as in Glenberg et al.'s (in press) experiment, memory for the texts was enhanced by observing the outcome of manipulation (effect sizes varied from 1.23 to 1.75 for different measures).

Richmond and Glenberg (2006) investigated an application of physical manipulation and imagined manipulation to the learning of an abstract principle in science, the *control of variables strategy* (e.g., Triona & Klahr, 2003). The principle of the control-of-variables strategy is that when designing an experiment, all variables other than the focal independent variable should be controlled, or held constant. In research with fourth-grade students, we are investigating whether initial physical manipulation helps the children implement an imagined manipulation strategy when later asked to learn from text. In the experiment, children have available the apparatus for setting up experiments to investigate questions such as whether the shape of an object affects the speed with which it sinks in a cylinder of water. In one condition, children read instructions and literally set up an experiment by controlling some variables (e.g., the height from which an object is dropped), while manipulating another variable, such as object shape. In another condition, children read the same instructions, but the experimenter sets up the experiment out of the children's sight. Later, all children are asked to judge the adequacy of other experiments from text alone. Whereas initial results are encouraging, we do not yet have enough data to draw definitive conclusions.

In summary, the point of reading is to convey meaning. But what is meaning? According to the IH, meaning arises from creating or simulating the perceptual/action situation described by sentences. These simulations are determined by the properties of the objects referred to, that is, the affordances of the objects, not the properties of the words. Physical and imagined manipulation help children to index words to objects so that affordances can be derived and meaning achieved. In chapter 12 of this volume, Johnson-Glenberg reviews other data demonstrating the importance of the application of visually based strategies for reading comprehension.

ACTION AND MATHEMATICAL STORY PROBLEM SOLVING

A good deal of evidence indicates that story problems are particularly difficult for children (e.g., Cooney & Swanson, 1990; Fuchs & Fuchs, 2002; Light & DeFries, 1995; Robinson, Menchetti, & Torgesen, 2002). We speculated that one reason for this difficulty is that some of the children were not adequately understanding the story, rather than their having some particular difficulty with math. If physical manipulation and imagined manipulation improve story understanding, will they also improve performance on story problems?

We created six types of problems set in two scenarios: (a) a fair (with rides, tokens, two children, a clown, and balloons) and (b) a zoo (with various animals, food for the animals, and a zookeeper). The problem types ranged from relatively simple addition problems to those requiring a combination of addition and multiplication (or repeated addition), to division. Figure 9.2 contains one of the problems calling for a combination of multiplication and addition. We refer to these as *Each problems* because of the importance of the word *each*; that is, in these problems, the word *each* signals that there are multiple alligators and multiple hippos and that each of the several individuals must be fed.

Children in the third grade were randomly assigned to one of three conditions. In the *story-relevant manipulation condition*, children manipulated the characters and objects much as in Glenberg et al.'s (2004) experiments. For example, in dealing with the Each problem illustrated in Figure 9.2, children would count out the appropriate number of little toy fish and distribute them to the animals. The *abstract manipulation* group was designed to simulate a type of manipulation that occurs in classrooms, such as the use of counting sticks to help solve problems. Note that these manipulables are often unrelated to the plot content of the story problem. Thus, children in the abstract-manipulation condition counted out Lego pieces whenever there was a green light sentence. These pieces were not shaped like the objects in the stories (e.g., fish or balloons); neither were the story objects present. On the other hand, the abstract manipulation condition does control for physical activity and any benefits of counting out objects. Children in the *reread* group read each sentence aloud and immediately reread those sentences with green lights. All children were told that the green light sentences contained particularly

Feeding the Hippos and the Alligators

There are 2 hippos and 2 alligators at the zoo.

They live by each other, so Pete the zookeeper feeds them at the same time.

It is time for Pete to feed the hippos and the alligators.

Pete gives each hippo 7 fish. 

Then he gives each alligator 4 fish. 

The hippos and alligators are happy now that they can eat.

How many fish do both the hippos and the alligators have altogether before they eat any?

Figure 9.2. Top: Example of an Each problem from the zoo scenario. Bottom: The zoo scenario toys.

important information. After reading the problems, children wrote an answer and a number sentence (or equation), indicating how they solved the problem. Because these two measures produced essentially identical findings, we focus on the answers measure.

Children participated in three sessions. On the first day, they solved problems related to one of the scenarios, and on the second day they solved problems related to the other scenario. On the third day, children in the story-relevant manipulation condition were taught imagined manipulation, children in the abstract-manipulation condition were asked to imagine counting out the Lego pieces, and children in the reread condition were asked to reread green light sentences silently. All children worked on problems from both scenarios during this third session.

The initial analyses of the results were quite disappointing in that there were no statistically significant differences among the story-relevant manipulation, abstract-manipulation, and reread groups. On closer inspection, however, we found an intriguing effect dependent on problem type. Some problems appeared to be too easy and resulted in close to ceiling effects; other problems (e.g., the division problem) were too difficult for most of the children. However, the data from the Each problems were just right: Collapsing across the first two days, the percentages correct were 75%, 28%, and 35% in the story-relevant manipulation, abstract-manipulation, and reread conditions, respectively (d = 1.09 for the contrast of story-relevant manipulation to abstract manipulation). The story-relevant manipulation advantage was also found on Day 3, when imagined manipulation was used: 50%, 40%, and 13% for the story-relevant manipulation, abstract-manipulation, and reread conditions, respectively. However, the differences on Day 3 were not statistically significant—perhaps because of low statistical power in that, on average, there were fewer than one Each problem per child on Day 3.

The second experiment was designed to follow up on these findings in several ways. First, given that the effect with Each problems was not predicted, we needed to replicate that finding. Second, there were only two Each problems (one for each of the scenarios), so the finding might have been due to idiosyncrasies with those two problems. In the second math experiment, there were six Each-type problems for each of the scenarios. Third, each child solved two Each-type problems on Day 3 using an imagined-manipulation strategy. The sessions were all exactly 1 week apart, so finding an effect of condition on Day 3 would indicate an effect of physical manipulation lasting at least 1 week.

Fourth, we tested a hypothesis as to why story-relevant manipulation would be particularly effective for Each problems but not others. Suppose that these third-grade children are not particularly skilled in understanding some grammatical constructions, such as the word *each*. Note that the correct understanding of *each* is quite complicated: The word must be indexed to the correct set of objects, and the objects within the set need to be individuated. For example, it is not sufficient when reading *each* for the reader to note that there is a group of alligators. In addition, the child must realize that there are two alligators and that both are fed individually. The story-relevant manipulation condition makes this individuation relatively obvious, however, because the child has to count out fish for each of the alligators (or, in the Fair scenario, count out tickets for each of the rides). In fact, the most frequent error in both the abstract-manipulation and reread groups was to count out only once, as if the child failed to individuate the animals in a group. For example, for the problem in Figure 9.2, the

TABLE 9.2
Proportion Correct Answers from the Second Math Experiment

	Sessions 1 and 2 (Physical Manipulation)		
Condition	*Each*	*Each Plural*	*Each Enumerated*
Story-Relevant Manipulation	.78	.72	.72
Abstract Manipulation	.50	.61	.67
Re-read	.33	.27	.40
	Session 3 (Imagined Manipulation)		
Condition	*Each*		*Each Enumerated*
Story-Relevant Manipulation	.76		.76
Abstract Manipulation	.38		.62
Re-read	.40		.47

most frequent answer in the abstract-manipulation and reread conditions was 11, rather than the correct answer of 22.

To test the individuation explanation, we manipulated the degree of linguistic support for the correct interpretation of *each*. In Each Plural problems, the wording of the green light sentences was changed from, for example, "Pete gives each hippo 7 fish" to "Pete gives each of the hippos 7 fish." That is, the noun was explicitly marked as plural. In Each Enumerated problems, the wording was changed to "Pete gives each of the 2 hippos 7 fish." That is, the noun was marked as plural and the actual number (2) was given. We reasoned that providing more linguistic support, so that children in the abstract-manipulation and reread conditions did not have to depend solely on their ability to index the word each, should improve performance and reduce the difference between these groups and the story-relevant manipulation group.

The data from the second experiment are presented in Table 9.2. Note that we successfully replicated the effect from the first experiment; namely, for Each problems, children in the story-relevant manipulation condition were far more successful than children in the abstract manipulation (d = 0.80) and reread conditions (d = 1.5). Also, the effect was maintained on the third, imagined manipulation, day; that is, children who had previously used physical manipulation solved problems more accurately when using imagination than children in the other conditions (d = 0.73).

Nonetheless, there was only modest support for the hypothesis that Each problems are difficult because children do not use the word *each* to index a group composed of individuals. First, the effect of problem type (Each, Each Plural, or Each Enumerated) was not statistically significant; neither was the interaction with condition statistically significant. Second, whereas there is a hint that adding linguistic support helped the children in the abstract-manipulation

TABLE 9.3
Problem Type 1 From the First Math Experiment
Feeding the Monkey and the Polar Bear

Pete the zookeeper is in charge of feeding animals at the zoo.
In the morning Pete feeds the polar bear and the monkey.
First he feeds the polar bear 9 fish.*
Then he feeds the monkey 5 bananas.*
In the afternoon he feeds the polar bear 8 fish.*
Now, Pete remembers he has to feed the monkey again.
If the monkey gets the same number of things as the polar bear, how many bananas should Pete give the monkey in the afternoon?

*Sentences followed by green lights.

group, linguistic support had mixed effects in the story-relevant manipulation and reread groups. In summary, although we have demonstrated some very large effects of manipulation and imagined manipulation in story problem solving, the boundary conditions for the effect are not yet clear.

We are currently investigating two new hypotheses in addition to the individuation hypothesis. The first new hypothesis is that, when faced with story problems, many children will ignore the story plot. When the plot provides information useful for problem solving (as in Each problems), these children will suffer. Because story-relevant manipulation forces the child to pay attention to the plot, children in the story-relevant manipulation condition outperform other children when the plot conveys important information.

A second new hypothesis addresses the congruence between the mathematical structure of the problem and the structure revealed by manipulation. For Each problems, the actions used in story-relevant manipulation mimic the structure of the mathematics; that is, the child counts out each number twice corresponding to the multiplication or multiple additions required to solve the problem. In contrast, consider the problem in Table 9.3. For this sort of problem, the counting out reveals some of the mathematical structure, but important steps are missing. That is, after counting out 9, 8, and 5, the solution requires the summing of 9 and 8 and then subtracting 5. The simple story-relevant manipulations do not mimic these latter steps. In fact, for two reasons, the story-relevant manipulation children might do poorly on the type of problem illustrated in Table 9.3. First, the manipulations leave the child hanging, because they do not specify the final steps. Second, it may be difficult for the child to conceptualize how to subtract 5 bananas from 17 fish! In fact, in the initial math experiment, story-relevant manipulation children were correct on only 46% of these problems, whereas children in the other conditions were correct on 60% of these problems. A similar analysis can hold for the division problems used in Experiment 1; that is, a simple counting-out strategy does not reveal much of the structure that corresponds to division (e.g., dividing the tokens into equal-sized piles). This hypothesis suggests that manipulation more sophisticated than simply counting out tokens may be more broadly applicable. For example, manipulation that emphasizes one-to-one correspondence might help children to solve problems such as those in Table 9.3, in which children must consider the equivalence between the number of fish and the number of bananas. Similarly, teaching manipulation in which equal-sized piles are created might help children to ground the concept of division. Nonetheless, these speculations remain to be tested.

IS MANIPULATION PRACTICAL FOR THE CLASSROOM?

Three aspects of our results suggest that manipulation may work well in the classroom. The first result is evidence for strategy maintenance and transfer. As described earlier, Glenberg et al. (2004) found that children who had brief training in physical manipulation followed by imagined manipulation maintained the strategies without further instruction several days after the last instructed session. A type of maintenance was also found in the second math experiment. Namely, the third session (in which imagined manipulation was taught) was a week after the last physical manipulation session. Nonetheless, the children were able to use memory of their physical experiences to enhance problem-solving performance compared with children who only read the problems. Finally, data from a third math experiment (Glenberg, Rischall, & Jaworski, in preparation) demonstrate both maintenance and transfer. That experiment was similar to the second math experiment reported here, except that 1 to 3 weeks after Session 3, the children were tested again in their classrooms with no scenario toys present. This test consisted of two Each-type problems, one from a scenario with which the child was familiar and one from a different scenario. For the familiar scenario, children who had had story-relevant manipulation (physical manipulation followed by imagined manipulation) solved 69% of the problems, whereas children who had reread solved only 46% ($d = 0.46$). For the unfamiliar scenario, the children never saw the corresponding toys; neither were they given any introduction to the scenario. Nonetheless, the children who had practiced story-relevant Manipulation earlier solved 46% of the problems compared to the reread children, who solved only 23% ($d = 0.48$). Thus, after fairly minimal training, the strategy is maintained for at least 3 weeks, and it transferred to a new scenario in a new environment.

A second reason for expecting success in the classroom comes from data demonstrating the effectiveness of group strategy training with children from diverse backgrounds. In Glenberg et al.'s (in press) study, low-income children attending after-school programs in community centers were given physical manipulation or reread training in groups of three. As noted earlier, the physical manipulation strategy was quite effective in this situation. Also, as Marley et al. (2005) reported, Native American children with learning disabilities benefited from listening combined with seeing the results of the experimenter's manipulations.

A third reason for optimism stems from the success of manipulation techniques in schools. As Lillard (2005) described, many of the successful methods used in Montessori schools take advantage of physical manipulation. In fact, Maria Montessori wrote:

> One of the greatest mistakes of our day is to think of movement by itself, as something apart from the higher functions … Mental development must be connected with movement and be dependent on it. It is vital that educational theory and practice should become informed by this idea. … Watching a child makes it obvious that the development of his mind comes about through his movements … Mind and movement are parts of the same entity. (Montessori, 1967, pp. 141–142)

Thus, Montessori agrees: Brains are for action!

ACKNOWLEDGMENTS

This work was supported by a grant from the Institute of Education Sciences to Arthur Glenberg and Joel Levin and University of Wisconsin Hilldale Awards to Beth Jaworski and Michal Rischall. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Institute of Education Sciences.

REFERENCES

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.

Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory & Cognition*, *32*, 863–873.

Casenhiser, D., and Goldberg, A. E. (2005). Fast mapping of a phrasal form and meaning. *Developmental Science*, *8*, 500–508.

Cooney, J. B., & Swanson, H. L. (1990). Individual differences in memory for mathematical story problems: Memory span and problem perception. *Journal of Educational Psychology*, *82*, 570–577.

De Vega, M. (in press). Levels of embodied meaning: From pointing to counterfactuals. In M. de Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning.* Oxford, England: Oxford University Press.

De Vega, M., Robertson, D. A., Glenberg, A. M., Kaschak, M. P., & Rinck, M. (2004). On doing two things at once: Temporal constraints on actions in language comprehension. *Memory & Cognition*, *32*, 1033–1043.

Fuchs, S., & Fuchs, D. (2002). Mathematical problem-solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities*, *35*, 563–573.

Gibson, J. J. (1979). *The ecological approach to visual perception*. New York: Houghton Mifflin.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*, 1–19.

Glenberg, A. M. (in press). Language and action. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics*. Oxford, England: Oxford University Press.

Glenberg, A. M., Brown, M., & Levin, J. R. (in press). Enhancing reading comprehension in small groups using a manipulation strategy. *Contemporary Educational Psychology*.

Glenberg, A. M., & Gallese, V. (2006). *The motor theory of syntax and semantics*. Manuscript in preparation.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Kaschak, M. P., & Japuntich, S. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, *96*, 424–436.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*, 558–565.

Glenberg, A. M., Rischal, M., & Jaworski, B. (in preparation.) *Contributions of a reading intervention to story-problem solving*.

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, *28*, 1–26.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*, 379–401.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301–307.

Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language*, *43*, 508–529.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Light, J. G., & DeFries, J. C. (1995). Comorbidity of reading and mathematical disabilities: Genetic and environmental etiologies. *Journal of Learning Disabilities*, *28*, 96–107.

Lillard, A. (2005). *Montessori, the science behind the genius*. Oxford, England: Oxford University Press.

Llinas, R. (2001). *The* i *of the vortex*. Cambridge, MA: MIT Press.

Marley, S. C., Levin, J. R., & Glenberg, A. M (2005, April). *Can text-relevant motor activity improve the recall of learning-disabled Native American children?* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Masur, E. F. (1997). Maternal labeling of novel and familiar objects: Implications for children's development of lexical constraints. *Journal of Child Language*, *24*, 427–439.

Montessori, M. (1967). *The absorbent mind* (C. A. Claremong, Trans.). New York: Holt.
Neininger, B., & Pulvermüller, F. (2003). Word-category specific deficits after lesions in the right hemisphere. *Neuropsychologia*, *41*, 53-70.
Rayner, K., Foorman, B. R., Perfetti, E., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest Monograph*, *2*, 31–74.
Richmond, E. K., & Glenberg, A. M. (2006). *Using concrete actions to understand an abstract principle*. Manuscript in preparation.
Robinson, C. S., Menchetti, B. M., & Torgesen, J. K. (2002). Toward a two-factor theory of one type of mathematics disabilities. *Learning Disabilities Research & Practice*, *17*, 81–89.
Pulvermüller, F. (in press). Grounding language in the brain. In M. de Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment and meaning*. Oxford, England: Oxford University Press.
Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, *74*, 209–253.
Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Triona, L. M., & Klahr, D. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition and Instruction*, *21*, 149–173.

10

Engagement Practices for Strategy Learning in Concept-Oriented Reading Instruction

John T. Guthrie, Ana Taboada, and Cassandra Shular Coddington
University of Maryland

The rationale for considering engagement in the instruction of reading strategies is presented. Processes of engagement include the use of prior knowledge during reading, the use of cognitive strategies, and motivations for reading comprehension. All three processes are shown to contribute independently to reading comprehension in empirical studies, and thus the role of motivation is suggested to be important. The experimental effects of engagement-supporting instruction on reading comprehension are presented, with an emphasis of practices of using conceptual goals, affording student choice, and providing opportunities for text-based collaboration. The potential theoretical explanation for these instructional effects is that the instruction increases students' levels of engaged reading, which then increases reading comprehension. An example of the principles embodied in a computer-based environment is provided.

Three questions with respect to the book's theme are addressed in this chapter. One theoretically based empirical study is presented for each question. The first question is: "Why consider engagement, and what are the processes of engagement?" Whereas the majority of the chapters in this book address the cognitive characteristics of strategies for reading comprehension or the assessment of cognitive competencies related to reading strategies, we propose here that motivation and engagement for reading are vitally important for strategy learning. In the opening section, we describe what we mean by *engagement* and give a brief snapshot of engagement's role in strategy learning and strategy use.

The second question is: "Does instruction for engagement improve comprehension strategies?" If indeed processes of engagement and motivation correlate with reading comprehension and strategies for comprehension, to what extent can these engagement processes be increased through instruction? If they are impervious to instruction, these processes are less relevant to the educational purposes of this book. However, if engagement in reading can be facilitated, then its role in contributing to strategies and comprehension becomes more interesting.

The third question is: "Why does instruction for engagement improve reading comprehension and reading strategies?" Here we address the mechanism by which instruction that facilitates engagement may also increase reading comprehension. In one sense, this is a purely scientific endeavor, which is intended to create a more thorough account of the role of engagement in strategy learning. In another sense, we believe that understanding the mechanism by which instruction affects comprehension through reading engagement can palpably improve instructional design.

WHY CONSIDER ENGAGEMENT, AND WHAT ARE THE PROCESSES OF ENGAGEMENT?

A substantial body of empirical research documents the close association of motivational variables and cognitive variables in reading comprehension. For students from elementary school through college, the use of strategies for reading comprehension correlates with students' competence in text comprehension (Pintrich, 2003). To understand complex, unfamiliar text, skilled comprehenders use such reading strategies as questioning and comprehension monitoring. More specifically, for college students, their use of strategies such as on-line summarizing, questioning, and deliberate attempts to integrate text are substantially associated with variables that can be described as internal motivation (Pintrich, 2003). In this context, by *internal motivation* we refer to students' desire to fully understand the material, their interest in processing the information deeply, and their enjoyment in learning the content. These internal motivations are contrasted with students' performance goals, which consist of finishing a task quickly, exerting minimal effort, taking shortcuts, and avoiding hard work (Elliot, 1999). For elementary and middle school students, as well as college students, the internal motivations are highly associated with competence in reading comprehension and the use of strategies in comprehension. In contrast, students' extrinsic motivations for learning tend to be uncorrelated or even negatively correlated with their use of high-level strategies for learning from text (Pintrich, 2000).

RATIONALE FOR STUDY 1: WHY CONSIDER ENGAGEMENT, AND WHAT ARE THE PROCESSES OF ENGAGEMENT?

Our engagement model of reading is partly based on this high association of internal motivation with strategies for reading comprehension. We define *engagement* as the joint functioning of the use of cognitive strategies, internal motivation, the use of prior knowledge for learning from text, and social collaboration during reading (Guthrie & Wigfield, 2000). In Figure 10.1, these engagement processes represent the mechanisms that are responsible for achievement and competence in reading comprehension. On the outside of the model are the environmental characteristics within classrooms or instructional programs that facilitate engagement and thereby increase reading comprehension. In other words, in our engagement model of reading development, a variable such as strategy instruction may increase strategy use (or other engagement processes), which consequently improves comprehension. The same can be said for other variables, such as autonomy support. That is, autonomy support may increase internal motivation (and possibly strategy use), which thereby increases reading comprehension.

At the top of Figure 10.1, we emphasize the importance of the teacher setting knowledge goals for reading instruction. The teacher uses this practice, as well as encourages students to set knowledge goals during reading. In the model, there is no logic to the positioning of the variables around the outside circle. We propose that each instructional variable contributes to engagement and that the relations among the instructional variables should be investigated. Next, we address each basic question of this chapter with an empirical study.

Research Questions for Study 1

In an in-progress investigation, Taboada and Guthrie (2006) posed the following question: "To what extent do the processes of background knowledge, reading strategies (questioning), and internal motivation contribute uniquely to reading comprehension?" Although previous studies have shown pairs of these variables to be correlated, they have not examined them in concert with

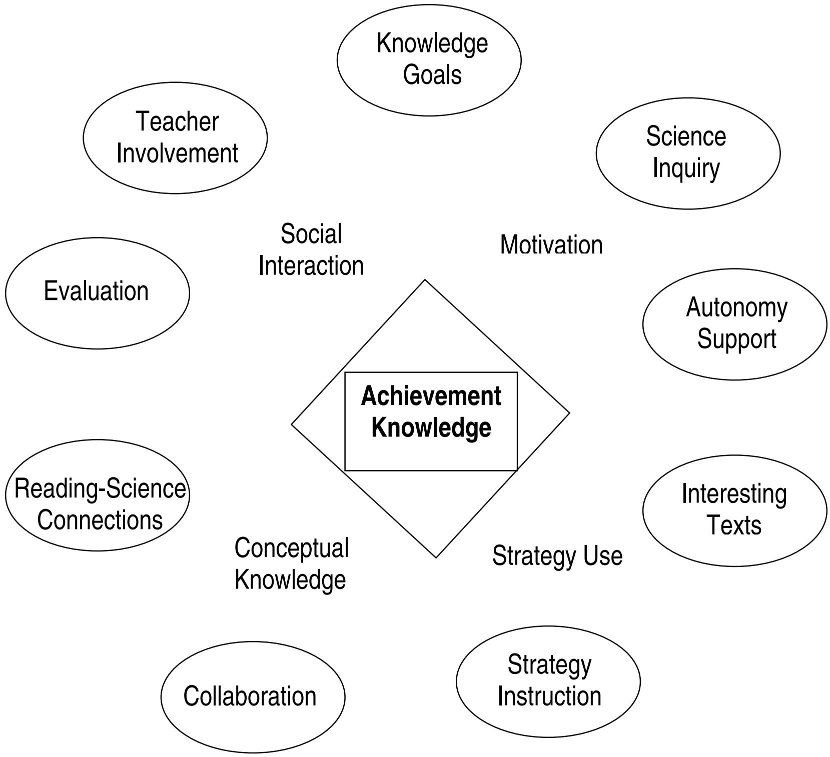


Figure 10.1. Engagement model of reading development.

each other. That is, reading strategies have been associated with comprehension, but we do not know whether that correlation remains when motivation is entered into the picture. Likewise, motivation has been associated with reading comprehension but may not be influencing comprehension when reading strategies are also considered.

Method for Study 1

To examine this question, we administered a variety of measures to 287 fourth-grade students from five schools in a small mid-Atlantic city. The students were approximately 50% males and 50% females, and were approximately 66% Caucasian, 20% African American, and 14% other minorities.

The measures, administered in December 2003, included the Gates–MacGinitie Reading Comprehension Test and a measure of multiple-text comprehension that required the students to select from among several texts to answer three broad comprehension questions and to write their new understanding at the conclusion of this reading activity. Writing was judged on a 6-level rubric, from a few simple facts (low) to an integrated pattern of relationships among major concepts (high).

The strategy measure consisted of an assessment of students' questioning. Students were given the opportunity to browse multiple texts for 2 minutes. Students then wrote up to 10 questions that would help them learn about the topic by reading those texts. The questions were evaluated on a 4-level rubric that progressed from a few simple facts (low) to the request for a pattern of explanatory concepts (high). (See Guthrie et al., 2004, for details of the rubric).

A measure of motivation consisted of a teacher rating of each student's internal motivation for reading on a scale of 1 through 4. Student were rated on the scale to the extent that they could be described by the following statements: (1) often reads independently, (2) reads favorite topics and authors, (3) is a confident reader, (4) thinks deeply about the content of texts, and (5) enjoys discussing books with peers. This rating scheme has high internal consistency (Cronbach's alpha exceeds .90).

Data Analysis for Study 1

We conducted three multiple regressions with the dependent variable of the Gates–MacGinitie Reading Comprehension Test and the independent variables of prior knowledge, questioning, and internal motivation. Table 10.1 shows that all of the independent variables uniquely correlated with the Gates–MacGinitie Reading Comprehension Test. As indicated in Table 10.1, prior knowledge had a beta of .23 ($p < .01$) with reading comprehension when strategies and motivation were statistically controlled. Questioning, our example of a strategy for this investigation, correlated uniquely with reading comprehension $\beta = .34$ ($p < .01$). Internal motivation correlated uniquely with reading comprehension at $\beta = .25$ ($p < .01$). As shown in Table 10.2, prior knowledge, questioning, and internal motivation also had unique correlations with multiple-text comprehension.

Conclusions From Study 1

From these data, we conclude that internal motivation for reading showed significant and unique contributions to reading comprehension, independently

TABLE 10.1
Predicting Gates-MacGinitie Reading Comprehension

Variable Status	*Variable Name*	R^2	R^2 *Cha.*	*Sig. Cha.*	β
Dependent variable	Gates-MacGinitie				
Independent variable	Prior knowledge	.17	.17	.01	.23*
Independent variable	Questioning	.29	.12	.01	.34*
Independent variable	Intrinsic motivation	.35	.06	.01	.25*

Note. Sig. = significant.

TABLE 10.2
Predicting Multiple Text Comprehension

Variable status	*Variable name*	R^2	R^2 *Cha.*	*Sig. Cha.*	β
Dependent variable	Multiple text comprehension				
Independent variable	Prior knowledge	.18	.18	.01	.31*
Independent variable	Questioning	.24	.06	.01	.25*
Independent variable	Intrinsic motivation	.26	.02	.01	.14*

Note. Sig. = significant.

of prior knowledge and questioning. In addition, it is obvious that questioning, one strategy for comprehension that was examined, significantly and uniquely correlated with reading comprehension, independently of internal motivation and prior knowledge. From these findings, we infer that these three processes of reading engagement appear to contribute to reading comprehension. This implies that motivation should be included explicitly in instruction that attempts to increase reading comprehension. It should be noted that this study was performed with Grade 4 students reading paper-based texts on a life science topic of ecology. Although we believe that the findings are generalizable to older students reading other topics in other environments, empirical studies are needed to confirm this expectation.

The results from the preceding investigation suggest that these processes of engagement, consisting of internal motivation, the use of background knowledge, and the use of strategies such as questioning for reading comprehension are all contributors to competence in reading comprehension. However, the processes are also correlated with each other to a moderate degree (about .5 for fourth-grade students), suggesting that it is legitimate to refer to engagement as a composite of cognitive and motivational processes that optimally predict comprehension. For instructional attempts to improve reading comprehension, we suggest that it is potentially valuable to address the full composite of engagement, including knowledge, strategies, and motivation in concert. We expect that the attempt to increase reading strategies directly as a medium for improving comprehension will also be facilitated by instructional attention to reading motivation.

DOES INSTRUCTION FOR ENGAGEMENT IN READING IMPROVE READING COMPREHENSION STRATEGIES?

Research Question for Study 2: To What Extent Does Concept-Oriented Reading Instruction Improve Reading Comprehension Strategies?

The second question addressed in this chapter is whether reading instruction that incorporates support for motivation increases students' use of reading strategies. At the same time, we inquired as to whether such engagement-supportive instruction also improves reading comprehension itself. In this section, we describe five practices for supporting engagement and motivation. We also discuss practices for scaffolding student engagement. We then present two recently published studies addressing the key question of this section.

Instructional Model for Improving Reading Motivation

Our model incorporated five instructional practices implemented by regular classroom teachers in Grades 3 through 5. The instructional models had either 12-week units or a 36-week units that used a plethora of information books and literary texts in an array of reading and writing activities with the five instructional practices at the heart of the instruction: (a) knowledge goals for reading instruction in a conceptual theme, (b) real-world interactions related to the knowledge goals, (c) student choice and self-direction in reading activities, (d) interesting texts for instruction, and (e) student collaboration in reading and writing. We portray each practice next.

Knowledge Goals for Reading Instruction in a Conceptual Theme. This practice is defined as teaching for understanding through conceptual themes about enduring and important concepts (Blumenfeld, 1992). In other words, a pre-eminent goal in instruction is to gain knowledge through information text or literary experience through narrative text. This practice contrasts with the frequent practice in reading instruction of placing a strategy as the pre-eminent goal and attaching topics only incidentally to reading strategies. In this circumstance, students are exposed to isolated topics, with constantly changing facts and principles. Under these conditions, deep knowledge cannot be acquired, and the transfer of strategies across texts is prohibitively difficult. Both of these barriers represent disadvantages for strategy instruction (Brown & Campione, 1998).

Instruction that integrates conceptual themes with reading instruction is called *Concept-Oriented Reading Instruction* (CORI), because concepts direct the teacher's and the students' goals for reading. In a recent Grade 5 implementation, the conceptual theme was "Plant and Animal Communities." The concepts of mutualism (plants and animals that support each other's survival) and other interactions, such as predation, were prominent. Also central were survival concepts of locomotion, feeding, competition, communication, reproduction, and others (Guthrie & Scafiddi, 2004).

It may seem counterintuitive to place knowledge goals as a motivational practice in reading. However, we believe it is motivating to learn about how animals survive; how sharks eat; how frogs freeze in the winter without dying; and, most of all, how the diversity of animals all acquire fundamental processes such as defense and locomotion to adapt to their habitats. Gaining this information through integrated and enduring themes gives students a sense of expertise that is highly rewarding for their reading pursuits. The effort of comprehending text is rewarded by new, exciting facts. More important, students are proud of the expertise they acquire by learning new concepts that can be explained and transferred to new domains.

The CORI practice of using knowledge goals in a conceptual theme is motivating, but it is also likely to support strategy development because it enables strategies to be used in a rich domain of interesting information. Therefore, strictly speaking, we claim that these practices are engagement supporting because they build internal motivations and foster cognitive strategies simultaneously. Obviously, the strategies may also be seen as motivational because they encourage and engender the development of internal motivations for reading activities.

Real-World Interactions. The instructional practice of using real-world interactions to teach reading consists of providing sensory experiences and multimedia experiences that are connected to reading activities (Sweet, Guthrie, & Ng, 1998). Our main reason for introducing real-world interactions is that they quickly establish a purpose for reading. When students interact with such life science artifacts as an owl pellet, an ant farm, a carnivorous plant, or a salamander, they become curious. They want to know more about the object of their observation. Spontaneously, they pose questions, bring up their prior knowledge, and seek information in books and on Web sites. Students' personal goals for reading are fostered by real-world interactions because they pose questions and develop topical interests that they experience as unique to them. In essence, real-world interactions energize reading activity and strategies for text interaction by piquing students' curiosities and desire for knowledge. If the practice of real-world interactions fails at this mission, it fails to be useful in the CORI model.

Supporting Student Choice and Self-Direction. This practice refers to providing students with meaningful and academically significant choices during instruction. In other words, during the teaching of reading strategies, a teacher may give students choices about a particular topic to read, a particular text to inspect, a particular strategy to use, or a particular way to express their new-found knowledge (Reeve, 2004). In a videotaped teacher's lesson, we recorded an outstanding teacher providing five meaningful choices during a 10-minute strategy lesson on comprehension monitoring. Thus, supporting students' self-direction through choice and autonomy support is fused intimately with the direct instruction of reading strategies in the context of deep text interaction.

This practice of autonomy support during instruction is contrasted with a frequent practice of providing opportunities such as DEAR time (Drop Everything And Read). Teachers who provide these practices often have a highly teacher-directed lesson for 30 minutes and then provide student choice of reading any book for 30 minutes. Although this may be valuable, it is not autonomy support during instruction but rather a choice apart from instruction. However, it is the immersion of choice into instruction that facilitates the development of self-regulated use of strategies for reading. (Autonomy-supporting classrooms are described more fully by Stefanou, Perencevich, & DiCintio, 2004).

Interesting Texts for Instruction. Information trade books and literary texts such as legends, stories, and chapter books for reading instruction are interesting by virtue of their appearance, readability, topic, and connection to classroom activities. An abundance of books matched to students' oral reading levels enables them to read with fluency, and thereby to focus on gaining conceptual knowledge and using strategies well. For example, in Grade 3, for 12 weeks of CORI, students read the following: 36 books (class sets), 9 books (team sets), and 6 books (individual selections), for a total of 51 books, including expository and literary texts. Book sets were selected according to goals for knowledge development, strategy instruction, and type of reading activity.

Reading a very high volume of trade books for language arts instruction is a necessary condition, in our opinion, for significant reading comprehension growth.

Support for Student Collaboration in Reading and Writing. The practice of providing collaboration support consists of enabling students to work together in partnerships and teams to gain knowledge and experience from text and use strategies effectively. In other words, the social interaction facilitated in CORI is not a mere diversion. It refers to active interpersonal interchange to gain deeper meaning from text. A few rules and roles are needed in the partnerships and teams to assure that the talk is accountable to the conceptual purposes for reading. However, discussing one's understanding of newfound information is rewarding and motivating to students of many ages (Brown & Campione, 1998). Consequently, harnessing this social disposition for learning from text has benefits for students' cognitive and motivational development.

Scaffolding for Engagement in Reading. The notion of scaffolding instruction applies to the development of engagement just as it applies to the development of cognitive competencies in reading. Just as an instructional program might include processes of scaffolding to enable students to learn strategies for reading comprehension such as questioning, scaffolding can be used to facilitate motivational development in reading. Effective teachers are likely to scaffold for engagement just as they do for cognitive strategies by providing modeling and then releasing responsibility to students. For example, in a typical reading situation, an engaged reader (a) selects a topic or subtopic relevant to a reading goal, (b) performs a task relevant to the teacher's directions, (c) selects a text, (d) uses one or more reading strategies, (e) adopts a social arrangement for reading, (f) allocates time to reading, and (g) decides how to represent the knowledge or experience gained from reading. In the case of high scaffolding, the teacher might give six of these actions to the students and allow the students to self-select one of them. For example, the teacher might provide the topic, task, text, time, social arrangement, and mode of expression (write a short paragraph), while allowing the student to choose the text from which he or she worked. In this case, the teacher performs 90% of the task, and the student performs 10% of the task. In the case of a lower scaffold, the teacher might give approximately half of the actions and the student supplies approximately half of the actions in a 50–50 arrangement. The teacher might supply the topic (solar system), the task (explain one planet aloud), and the time (20 minutes). She would allow the students to select the text (a nonfiction book), the reading strategy (question, read, and discuss), the social arrangement (pairs or not), and the mode of self-expression (paragraph or drawing). With a minimal level of scaffolding, the teacher places responsibility for self-initiation in the hands of students; the teacher and students would participate in a 10%–90% arrangement. For example, a teacher might supply one action, such as providing a text (a new trade book) for the day's learning, and the students would supply the rest of the processes, such as choosing the topic within it, identifying a task to represent comprehension of their topic, allocating time, choosing a social arrangement, deciding a mode of expression, and selecting strategies for reading.

Teacher monitoring and feedback during these self-directed activities becomes vitally important. Students may not initially make good selections of the text or the task to be performed. However, a well-devised scaffolding process will assure that students are not lost or ill guided. This kind of scaffolding for self-directed reading would be expected to be shifting from high teacher direction and high scaffolding to student self-direction over a process of 3 to 4 months (October–February) in a normally functioning classroom. Student learning of this self-regulation would certainly not occur in a matter of hours or days, assuming that the goal is to have them become independent, productive learners in a classroom situation.

Scaffolding for Knowledge Goals in Reading Instruction. Just as students need to be self-regulating in the processes of engaged reading, teachers can provide scaffolding for the use of knowledge goals in reading comprehension activities. For example, when teachers are highly directive (90%–10% arrangement), they may supply the theme (communities of plants and animals), the topic (mutualism between two animals), and an example (the bird and the buffalo), and ask the students how the bird and the buffalo both benefit from their relationship. The students' responsibility is to read a page that describes the bird eating bugs from the buffalo's back (the bird gains a meal while the buffalo is freed from parasites). In a highly released situation with a low scaffold (10–90 relationship), the teacher may only provide the theme by saying, "Continue your work on plant/animal communities" and the students would choose a topic (e.g., predation) and several examples of the topic (a variety of predatory animals), and an illustration of survival processes for these animals and their prey. Thus, the teacher provides a scaffold for the process of constructing knowledge goals, creating subgoals, accomplishing the knowledge goals, and completing knowledge learning tasks.

Scaffolding for Collaborating and Using Interesting Texts in Instruction. Engaged learners collaborate effectively with each other. They locate texts that are interesting to them and that will help them gain knowledge and improve strategic reading. For example, a teacher may highly scaffold the collaboration process (90–10 relationship) by predetermining teams of students, roles of individuals within the teams, tasks to be performed by teams, and texts to be used. The student contribution consists of selecting the actual turn-taking activities within the team, including who will contribute, when, and how. In a low scaffold for collaborating, teachers may provide the announcement of teams and a global task for them to perform. Students would then select the roles within teams, the texts to be used, and the turn-taking procedure to be used in the team, as well as the time allocated to specific subtasks. Of course, this is a relatively mature form of self-guided collaborating, but effective teachers can readily scaffold it into third- to fifth-grade classrooms.

Likewise, the process of selecting interesting texts for learning and knowledge development can be scaffolded. As previously described, engaged readers are willing and able to locate texts that will help them gain both knowledge and strategies for reading. In a highly directed lesson, teachers may provide the book page and paragraph to be read and discuss how it is interesting with respect to the topic and, perhaps, its vivid details. The students may choose particular facts of importance within the text for emphasis or discussion. In a lower scaffold scenario, teachers may provide the guideline for a broad project in which students are learning about a content domain and the students would select texts, sections within them, and paragraphs or portions of high importance to the topic.

Studies 2A and 2B

Method and Design. In the next two studies, we used an analysis by synthesis approach that consisted of a hierarchical instructional design, as shown in Table 10.3. A control group in this design consisted of traditional instruction (TI) that had a substantial amount of interaction with text. A treatment group within this design consisted of strategy instruction (SI) that included text interaction and added the component of explicit SI. The third treatment condition in this design was CORI that included the components of text interaction and SI but also added motivational practices. By comparing these three conditions, we inferred the roles for SI and for motivational practices in classroom applications of reading comprehension instruction. (Further rationale for this analysis by synthesis approach is provided by Guthrie et al., 2004). For these studies, the control condition of TI included a basal program reading.

The condition of SI included basal programs supplemented by trade books with explicit teaching for the following strategies: activating background knowledge, questioning, searching for information, summarizing, organizing

TABLE 10.3
Analysis by Synthesis: Hierarchical Instructional Design

Motivational practices		
Strategy instruction	Strategy instruction	
Text interaction	Text interaction	Text interaction
CORI	Strategy instruction	Traditional instruction

graphically, and structuring stories. One week of explicit SI was provided for each of these strategies with a design for their integration for 6 weeks. SI consisted of explicit teaching beginning with modeling. The teacher showed how she would use a strategy, such as questioning, to help her learn key concepts from text. Then students performed the strategy with feedback for several days on multiple texts. As the teacher reduced the scaffold, she encouraged students to use high levels of the strategy (e.g., complex questions), as frequently as necessary to assure full understanding of texts. Later in the 12-week unit, teachers modeled and scaffolded the use of multiple strategies in an integrated sequence.

The condition of CORI consisted of the same strategies included in the SI, with the addition of the five practices for motivation support described previously consisting of knowledge goals for reading, real world interactions, autonomy support, interesting texts, and social collaboration.

The design was a pretest–posttest equivalent-groups plan, with 260 third-graders in four schools. Instruction occurred for 100 or more minutes daily for 12 weeks. Professional development was provided for 10 days during the summer and 5 days in follow-up monitoring activities.

Measures for Study 2A. The measures consisted of pretest and posttest administrations of multiple-text comprehension, passage comprehension, strategies (activating background knowledge, questioning, searching for information), and motivation for reading. In a performance assessment administered on 3 consecutive days in 60-minute sessions, students first wrote their background knowledge on the ecological topic of either (a) ponds–deserts, (b) oceans–forests, or (c) rivers–grasslands. Students' written statements were coded to a 6-level rubric. Next, students wrote up to 10 questions on their ecology topic. Each question was coded to a rubric, and the mean quality was determined (Taboada & Guthrie, 2006). For 40 minutes, students searched and read a 70-page packet of various reading levels on their topic. They recorded the sections from which they read and took notes. The number of relevant selections chosen (20% of the available sections were irrelevant) was the measure of search quality. Students then wrote an essay on their knowledge of the topic. Essays were coded to the same 6-level rubric used for prior knowledge to represent multiple text comprehension. Passage comprehension was a computer-based measure that used Pathfinder to determine the student's "concept map" learned from reading a specific 500-word text. The motivation measure was a student self-report of intrinsic motivation and self-efficacy for reading.

Results of Study 2A. Results from analyses of variance showed that CORI students and SI students were not different in the pretest and that CORI students had an advantage on the posttest that was statistically significant for passage comprehension (effect size = 1.32). On the measure of multiple-text comprehension, the groups were not different in the pretest, and CORI students had a significant advantage over SI students in the posttest with an effect size of 1.01. On the strategy composite, consisting of the combination of activating background knowledge, organizing information, and searching for information, the CORI and SI groups were not different in the pretest. However, CORI students had an advantage over SI students in the posttest that was statistically significant, with an effect size of 1.23. On a motivation composite, both CORI and SI groups were the same on the pretest, and the CORI group was significantly higher on the posttest with an effect size of 0.98. Therefore, CORI students surpassed strategy instruction students in passage comprehension, a reading strategy composite, and reading motivation.

Methods and Results in Study 2B

In Study 2B, we added a standardized reading test (Gates–MacGinitie Comprehension Reading Test), a new teacher-rated motivation measure (the Reading Engagement Index [REI]), and additional classrooms for all of the instructional groups. The same measures from Study 2A were administered in addition to the aforementioned new ones. The results showed that on passage comprehension, CORI students were significantly higher than SI students (effect size = 1.48). CORI students were also significantly higher than TI (effect size = 2.75). On the Gates-MacGinitie Comprehension Reading Text, CORI students were significantly higher than SI students (effect size = 1.40), and CORI students were significantly higher than TI students (CORI = .71). On motivation, CORI students were significantly higher than SI students on intrinsic motivation (effect size = 1.23), extrinsic motivation (effect size = 1.29), and self-efficacy for reading (effect size = 0.95).

Conclusions of Study 2. Our conclusions from Studies 2A and 2B are that CORI students surpassed both SI and TI students on passage comprehension. CORI students exceeded both SI and TI students on the Gates–MacGinitie

Reading Comprehension Test, and the CORI group was higher than the SI group on reading motivations measured with the REI. Our inference from these data is that motivational practices of CORI contributed to students' comprehension strategies and motivation, which is consistent with other findings (Taylor, Pearson, Clark, & Walpole, 2000). We note that it may be the practices alone, or motivational practices interacting with strategy instruction, that increase reading comprehension. We cannot attribute the effects solely to motivation practices with this design. However, without the motivation practices, it is evident that students' progress in comprehension with stragegy instruction alone was not remarkable and rarely different than traditional instruction. In other words, when motivational support is combined with SI and text interaction, CORI provides a value-added effect for reading outcomes compared with SI or TI groups. This value-added effect may be due to all or some interactions among motivational practices within CORI. We believe that these motivational practices are synergistic, although at present we lack empirical evidence on this point. We further believe the motivational practices are synergistic with strategy instruction, although we lack empirical evidence of those interactions at present.

WHY DOES INSTRUCTION FOR ENGAGEMENT WORK?

Rationale for Study 3: Why Does Instruction for Engagement Improve Reading Comprehension and Reading Strategies?

To this point, we have suggested that three processes of engaged reading facilitate reading comprehension: (a) use of background knowledge, (b) use of strategies for comprehending text, and (c) internal motivation for reading. We presented preliminary evidence that all three of these processes provide unique contributions to students' reading comprehension levels. As these processes each appear to be valuable for proficiency in comprehension, we suggest that it is potentially useful to facilitate all of them through instruction.

In the previous section, we showed that instruction that incorporates support for motivation combined with reading strategies increases comprehension of students in later elementary grades. Now the question becomes "What might account for the instructional effects on students' reading comprehension and reading strategies?" It is plausible that the instruction described previously simply increased the students' amount of reading, which then improved their comprehension. Equally possible is that this instruction increased the students' competence in collaborating, which improved their comprehension. We believe, however, that it was the extent to which students' reading was engaged and active that determined their growth in comprehension. We propose that engagement in reading as a student characteristic mediates the effect of instruction on the outcomes of reading comprehension. We further suggest that reading engagement mediates the effect of instruction on the development of the reading strategies themselves. In other words, students gain in generalized strategic competence to the degree that their reading activities during instruction are highly engaged.

Research Question for Study 3

The question driving Study 3 was "To what extent are the effects of reading instruction on reading comprehension mediated by students' levels of engagement in reading?"

Method of Investigation for Study 3

We conducted Study 3 with 300 fourth-grade students from 15 classrooms. Before and after a 12-week intervention, students were given assessment measures of reading comprehension (Gates-MacGinitie and a multiple-text comprehension task) and reading strategies. The reading strategies included questioning, activating background knowledge, and searching for information. Instructional variation consisted of classroom assignments to CORI, SI, and TI.

The two measures of reading engagement during instruction were a teacher rating (REI) and student portfolios. Using the REI, teachers provided an assessment of the extent to which each student was engaged in reading during the 12-week instructional period. Teachers rated each student on seven items: (a) often reads independently (motivation), (b) reads favorite topics and authors (motivation), (c) gets distracted while reading (motivation—reversed), (d) is a confident reader (self-efficacy), (e) uses strategies during reading (strategy use), (f) thinks deeply about the content of texts (knowledge use), and (g) enjoys discussing books with peers (collaboration). These were reliable ratings (Cronbach's alpha exceeds .90), and this measure correlates significantly with the Gates-MacGinitie Comprehension Reading Test ($r > .60$). The substantial alpha shows that teachers' ratings of students' strategies for reading comprehension were associated with their ratings of students' motivation, knowledge use, and collaboration. This suggests that teachers' perceived student engagement is a relatively well-integrated construct.

The second measure of reading engagement was based on students' text-based writing. During instruction, students were expected to complete portfolios that consisted of writing tasks connected to reading activities such as activating background knowledge, composing questions, recording their search for information, and composing a synthesis of how animals survive in plant/animal communities. The portfolios were coded on a rubric that reflected quantity and quality of engaged reading in classroom work (Guthrie et al., 2006).

Coding Rubric for the Drawings. We gave each science drawing a feature score ranging from 1 to 3, based on the extent of the labeling of the drawing's features. Each drawing also received a quality score that ranged from –1 to 1, which was based on the accuracy of the representation of the object being drawn. The feature and the quality scores for each drawing were summed to create a single final score, ranging from 0 to 4, which we recoded to range from 1 to 5.

Question Rubric. Each science process question was coded on a four-level questioning rubric (Taboada & Guthrie, in press). We coded simple science process questions that requested factual or trivial information at Level 1. Level 2 questions requested global information about a general ecological concept or an aspect of survival. Questions that probed for a more elaborated explanation about a specific ecological concept and contained evidence of an animal survival trait or characteristic were scored as Level 3 questions. Level 4 questions inquired about the complex interactions among multiple survival concepts or across multiple organisms.

Hypotheses Rubric. We coded all hypotheses in terms of the presence or absence of two components: (a) identification of the variables, and (b) description of the expected response of the organism or the object to the variable. For example, we gave a score of 1 to a hypothesis that simplistically described the change that the student expected to occur in the organism, without giving an explanation. We gave a score of 4 to an elaborated description of the expected effect with a reason consisting of the core survival concepts taught in the CORI classroom.

Rubric for Tables and Graphs. We coded each table and graph in the students' portfolios on a 2-point scale. Based on the analysis of the teacher-given or task-specific instructions, we gave a score of 2 to a table or a graph that appeared to be complete. Incomplete tables, those that were missing data, and incomplete graphs, those that were missing elements, received a score of 1.

Conclusion Rubric. Students' conclusions were coded on a scale in terms of four components: (a) a qualitative or quantitative representation of the experiment's results (1–3), (b) an explanation of results by means of variables

TABLE 10.4
Analysis of Variance/Covariance for Mediated Effects of Instruction on Comprehension

Dependent Variable	*Independent Variable*	*Covariate/Reading Engagement*	*F*	*Sig.*
Reading comprehension Gates-MacGinitie	Instruction	None	4.83	.029
Reading comprehension Gates-MacGinitie	Instruction	Teacher ratings	1.12	ns
Reading comprehension Gates-MacGinitie	Instruction	Students' portfolios	< 1	ns

(1–4), (c) a description of a cause-and-effect relationship between the variables (1–2), and (d) an elaboration of any aforementioned component (1–2). Each conclusion received a total score that resulted from totaling the scores from each of the four components. Conclusion scores ranged from a minimum of 4 to a maximum of 11. Total portfolio scores were the sum of the separate rubric scores.

Data Analysis and Results for Study 3

We analyzed the data to determine the extent to which students' engagement mediated the effect of instruction on reading comprehension. The three instructional conditions (CORI, SI, and TI) accounted for a significant amount of variance in reading comprehension on the Gates–MacGinitie Test and the multiple-text comprehension task. In Study 3, the teacher ratings (REI) correlated significantly with reading comprehension, and the portfolio measure correlated significantly with reading comprehension. To investigate this model, we conducted an analysis of variance to determine the level of instructional effect on reading comprehension, and we added a covariate of student engagement. If engagement mediates the effect of instruction on comprehension, the covariate removes or substantially reduces the effect of instruction on comprehension. As Table 10.4 shows, when reading comprehension on the Gates–MacGinitie was the dependent variable, instruction without a covariate had a significant effect on the reading comprehension outcome, $F(\text{xx}) = 4.83$, $p < .029$. Also, as Table 10.4 shows, when reading comprehension on the Gates–MacGinitie was the dependent variable and instruction was the independent variable, the measure of student engagement consisting of the REI entered as a covariate removed the instructional effect on reading comprehension. Furthermore, the same effect was found for the other measure of student engagement. When instruction was the independent variable, the measure of reading engagement consisting of student portfolio entered as the covariate removed the effect of instruction on reading comprehension on the Gates–MacGinitie. Therefore, it appears that the effect of instruction on reading comprehension was significant when engagement was not included in the analysis but was not significant when engagement was included as a covariate according to two different measures of student reading engagement. This leads to the possible conclusion that reading engagement mediated the effect of instruction on student reading comprehension. In other words, this interpretation is that the instruction that was designed to increase engagement in fact improved engagement in the form of high amounts of motivated, strategic reading. Then, student engagement increased student reading comprehension. Another interpretation of these statistical results is that the students in the CORI condition entered the study higher in engaged reading than students in the SI or TI conditions. However, this is unlikely, because according to a self-report, all students entered Study 3 equal on reading comprehension (Gates–MacGinitie) and motivation. However, in the absence of pretests on the two measures of engagement, which were not available, this latter interpretation cannot be ruled out.

Our original question asked whether student engagement accounted for the instructional effect on students' use of strategies. We analyzed the effect of instruction on reading strategies and the possible mediation by student engagement. As shown in Table 10.5, when a composite of strategies consisting of questioning and activating background knowledge was the dependent variable, and instruction was the independent variable, the significant effect of instruction was observed, $F(xx) = 6.03$, $p < .01$. However, when the composite of strategies was the dependent variable, instruction was the independent variable, and the measure of student reading engagement consisting of the REI was used as the covariate, the effect of instruction was reduced insignificance. Also, as shown in Table 10.5, when the composite of strategies was the dependent variable and instruction the independent variable and the measure of engagement consisted of the student portfolios was used as the covariate, the effect of instruction was also reduced to insignificance. From these findings, we conclude that the extent to which students were engaged in reading, according to teacher ratings or student portfolios, mediated the effect of instruction on the acquisition of reading strategies that consisted of activating background knowledge and questioning.

It appears that although instruction has positive effects on reading comprehension outcomes and the development of reading strategies, the effects are explainable in terms of the degree of student reading engagement. This assumes that student engagement was influenced by the instruction, and the data show that their engagement was indeed correlated with instruction.

TABLE 10.5
Analysis of Variance/Covariance for Mediated Effects of Instruction on Reading Strategies

Dependent variable	*Independent variable*	*Covariate/Reading engagement*	*F*	*Sig.*
Strategies	Instruction	None	6.03	.01
Strategies	Instruction	Teacher ratings	1.51	ns
Strategies	Instruction	Students' portfolios	< 1	ns

However, we do not have a pretest measure and a posttest measure of reading engagement. Consequently, we do not have a measure of the extent to which engagement changed during instruction. As a result, we cannot make the stronger claim that changes in student engagement were associated with changes in reading comprehension. We are limited to the more cautious claim that instructional variations had impacts on growth of reading comprehension and reading strategies to the extent that the instruction was correlated with students' levels of reading engagement within the classroom. In the end, we believe these data are highly suggestive for the proposition that student reading engagement is a mediating link between instruction designed to foster motivation and improvements in reading comprehension and use of reading comprehension strategies.

EXAMPLES OF MOTIVATION SUPPORT IN READING INSTRUCTION

The next section contains two illustrations of motivation support fused with strategy instruction for reading comprehension. First, we view Margie, a CORI teacher, providing instruction for fifth-grade students. This is a transcript of an actual videotape of raw classroom footage. It is an authentic lesson typical of Margie's teaching. Second, we present a hypothetical lesson in a computer-based learning environment. Although we have not observed such incorporation of motivation support within computer-based learning environments, we believe it is feasible and that design principles could be induced from this illustration.

Example of CORI Instruction

Margie has been using the CORI model for 6 weeks in her fifth-grade classroom. Her instruction is centered on the theme of plant and animal interactions and survival in communities. Today, her reading comprehension lesson is to teach students to integrate the two strategies of searching for information and questioning. To teach the strategy of searching, Margie will model and scaffold for her students the following four processes: (a) form a goal (through topic statement or question), (b) select text relevant to the goal, (c) read carefully in the selected text to identify goal-relevant information, and (d) integrate information from the text with prior knowledge and previously read information. These steps are placed on a chart near the students who are working to learn these strategies. Margie provides her students with choices during the course of this 15-minute reading activity. She sits with five students clustered at a small table at the side of the classroom. The table beside them has 10 to 15 colorfully illustrated texts on a variety of biomes. Margie begins the lesson.

> Sometimes when we read a text for new information, we skim. We are looking for new ideas to answers our questions. When we skim, we sometimes miss information. When we are searching for information, we are looking for answers to our questions, right? And we're using the text features to help us find information, right? But sometimes we forget to do this third step, which is to read carefully.

As she speaks, Margie points to a poster on the bulletin board entitled "Steps for Searching." The poster aids students through the process of searching information texts and helps them find answers to their own questions. Margie tells students to choose a text from those on the table. She instructs them to find information they have previously identified for their personal project and to choose a page for today's lesson on careful reading. Students select a text (trade book) and begin to search for relevant information. Margie then guides the students to place a removable adhesive note on a page where they already knew information before reading.

Next, Margie asks the students to reread the page very carefully and to think about every sentence. When the students have finished rereading, Margie asks them to place a new adhesive note on information they failed to notice the first time, but now realized, "Wow! I didn't know that!" Each student spends several minutes applying an adhesive note to a sentence that provides new information.

After the students have selected their sentences, Margie continues by asking the students to take the sentence they just found and write a question about some new information they want to find. Students spend several minutes on this question-writing activity. She then asks students to share the information they learned with the other group members. Michael begins by reading the sentence he chose, followed by his question: "Deserts are biomes that receive less than 10 inches of rainfall each year. I'm thinking, do they get any more rain than just 10 inches? How would all of those plants get to grow if they didn't have anymore rainfall?"

Margie asks the other students to respond to Michael's question, and they engage in a lively exchange about how cacti conserve water and other examples. Responding to Margie's request for a volunteer, Allison eagerly reads her question aloud to the group: "Does the rainforest ever run out, like for the meat eaters? Does the rainforest ever run out of food for them to eat?" Samantha responds immediately to Allison's question about the ecological balance of rainforests by saying, "That would probably be a thing we have to discover years from now." Everyone chuckles. Margie says, "You're going to have me thinking for the next 10 years! That's a nice question, Allison," and she proceeds to ask for more questions.

This guided reading group activity illustrates the use of autonomy support in the classroom. The lesson was clearly guided practice for using the combined strategies of searching for information and forming new questions during reading. Margie provided at least five meaningful choices during this 10-minute lesson segment. The first choice for students was to select a biome, which they had previously chosen to learn about (rainforests, deserts, grasslands, temperate forests, or arctic). Her students' second choice was to select the information book on their biome and the page in their book for today's lesson. On their teacher's request, the students' third choice was to record knowledge they previously possessed with respect to their selected page of text. Needless to say, students were at liberty to choose different information facts, and many students provided unique information at this time of the lesson. The students' fourth choice was identifying information that was new to them on the page. It is clear that information that is new to one student may not be new to another student, and a significant amount of latitude within the task structure was provided. Their fifth choice was for the students to pose their own questions based on their text choice. These questions will serve as the topical goals for the next reading activity.

Because students were working in different biomes, their questions were highly distinctive. In the videotape, students' enthusiasm and pride in their personal questions was self-evident. Students' curiosity for gaining new knowledge from text is visible in the reading activity but, more important, students' curiosity is explicitly supported through instructional actions. Fused with the strategy instruction, these actions generated energy, enthusiasm, and depth of conceptual thinking. Note the last question raised by the student, Allison: "Does the rainforest ever run out of meat for predators?" This refers to sustainability of habitat for life. This question is also adopted by the Sierra Club, advocated by the World Wildlife Fund, and addressed by the Nature Conservancy. Such a question represents high-caliber thinking from an otherwise-typical student in a Grade 5 classroom. We believe this depth of questioning is a consequence of internal motivation for reading, as well as cognitive competence in forming questions.

Example for Computer-Based Instruction for High School Students

Suppose we were designing and implementing a computer-based instruction to support students' strategy learning, such as iSTART or other electronic environments (see, e.g., chap. 16, this volume). How would motivation support be provided? Let us assume that the goals of this computer environment are to teach students reading strategies useful for information text that will enable them to comprehend more fully. In this imagined computer environment, a variety of texts are placed within several themes. The themes consist of such topics as "Ancient Greece" and "Threatened Ecosystems." Within the theme of Ancient Greece, six texts are available on the subjects of early democracy, commerce, the oracles, wartime, theater, and the gods. Within threatened ecosystems, six texts cover the topics of global warming, accelerated evolution, alien species, frozen frogs, human invasion, and rainforests. Before reading, the student selects a theme (Ancient Greece or Threatened Ecosystems) and, within the theme, they select one text from the six available.

Students' first action in the computer environment is to relate their personal prior knowledge to their selected text. Students are given an extremely simple quiz with a question and two alternative answers on which their success rate will be extremely high. Next, they view a brief video on the topic of their text. These activities activate their existing knowledge and link the topic to their background information. The intention of this computer program is to teach students the strategy of identifying key concepts that will be important to the macrostructure of the text. After viewing the video, students read a 500-word passage from their selected text. They are then instructed to select three words they believe are central and highly important to the meaning of the text. Students are informed that the passage has 10 important words, and their goal is to identify three. The computer program provides the full list of words and feedback on whether their three selected words were on the list.

Next, students set a goal for their subsequent activity. They decide how many key words they will attempt to identify on the next text. They are encouraged to choose a number that is a slight increase over the correct number they had identified previously. A student who correctly identified three words might be expected to set the goal of identifying four words in the next text. After reading another 500-word passage from a different text, and attempting to identify their selected number of words, students are provided feedback about the accuracy of their choices. Students repeat this activity for all six texts within their theme. The culminating activity for each text is for students to work, in pairs, to collaborate in forming a concept map of important key terms from the 500-word passage to represent their macrostructure of that text. Feedback provides the students with an ideal macrostructure of those terms for the text they chose.

This fictional computer-based environment for teaching students macrostructure had several motivation support for students, including internal motivations to comprehend, and self-efficacy in reading. Incorporated also into this example were the five design principles of CORI:

1. Use knowledge goals, by placing each text in a broader theme and culminating the instruction with the task of making an across-text concept map. In this case, the specific texts were contextualized in the themes of Ancient Greece or Threatened Ecosystems.
2. Provide real-world interaction with the topic, as nearly as possible. This consisted of a brief video clip relevant to the text.
3. Permit students to choose the texts they read, to identify key words they perceived as highly important. In this case, students selected which text within their chosen theme they would read first. Note that all students read all texts, but the order was their choice.
4. Use interesting texts with vivid details and visual appeal.
5. Arrange for student collaboration with feedback on its effectiveness.

With these five design principles functioning, students' internal motivation (interest, enjoyment, commitment to learning) is facilitated. In addition, students' self-efficacy (belief in their capacity to learn and to improve) is supported and potentially increased. Of course, this example is fictional. However, studies of motivational support in computer-based environments (Cordova & Lepper, 1996) have documented the impact of the conditions of personal connections and choice on perseverance, time spent, and interest in continued learning. They, therefore, merit experimental investigations.

SUMMARY

In this chapter, we have suggested that motivational processes facilitate the learning of skilled reading comprehension. As indicated in Study 1, internal motivations (desire to understand, enjoyment of reading, effort toward comprehension) made a unique contribution to students' competency in comprehension along with the contributions of background knowledge and reading strategies such as questioning. In view of the potential importance of motivational processes and comprehension, we have examined whether motivation can be increased with instruction. Our experiments reported in Studies 2A and 2B of this chapter and other research show that internal motivation is increased by a limited set of definable instructional conditions. In our framework, these conditions include conceptual goals, real-world interactions, autonomy support, interesting texts, and collaboration support. The extent to which each of these operates independently is unknown at present. We further suggested that these instructional characteristics increase reading comprehension because they enhance students' reading engagement; that is, the instructional effects are attributable to the level of engaged reading within the instructional activity. Highly engaged readers are motivated, strategic, and knowledgeable in their interactions with text, and their gains from instruction are higher than the gains for less engaged readers. We provided some commonplace examples of instruction infused with motivation support, one from a fifth-grade classroom and one from an imagined, computer-based learning environment. We suggest that although strategy learning is highly beneficial for reading comprehension, motivational development is also vital. Because motivation for reading is important, educational designers are likely to optimize their effectiveness by planning as explicitly for motivation as they do for students' cognitive functioning during instruction.

REFERENCES

Blumenfeld, P. C. (1992). Classroom learning and motivation: Clarifying and expanding goal theory. *Journal of Educational Psychology*, *84*, 272–281.

Brown, A. L., & Campione, J. C. (1998). Designing a community of young learners: Theoretical and practical lessons. In N. M. Lambert & McCombs (Eds.), *How students learn: Reforming schools through learner-centered education* (pp. 153–186). Washington, DC: American Psychological Association.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, *88*, 715–730.

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, *34*, 169–189.

Guthrie, J. T., & Scafiddi, N. T. (2004). Reading comprehension for information text: Theoretical meanings, developmental patterns, and benchmarks for instruction. In J. T. Guthrie, A. Wigfield, & K. C. Perencevich (Eds.), *Motivating reading comprehension: Concept-Oriented Reading Instruction* (pp. 225–248). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 403–424). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., et al. (2004). Increasing reading comprehension and engagement through Concept-Oriented Reading Instruction. *Journal of Educational Psychology*, *96*, 403–423.

Guthrie, J. T., Wigfield, A., Humenick, N. M., Perencevich, K. C., Taboada, A., & Barbosa, P. (2006). Influences of stimulating tasks on reading motivation and comprehension. *Journal of Educational Research, 99*, 232–345.

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, *92*, 544–555.

Pintrich, P. R. (2003). Motivation and classroom learning. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (pp. 103–122). New York: Wiley.

Reeve, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation & Emotion*, *28*, 147–169.

Stefanou, C. R., Perencevich, K. C., & DiCintio, M. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision-making and ownership. *Educational Psychologist*, *39*, 97–110.

Sweet, A. P., Guthrie, J. T., & Ng, M. (1998). Teacher perceptions and student reading motivation. *Journal of Educational Psychology*, *90*, 210–224.

Taboada, A., & Guthrie, J. T. (2006). Contributions of student questioning and prior knowledge to construction of knowledge from reading information text. *Journal of Literacy Research, 38*, 1–35.

Taylor, B. M., Pearson P. D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. *Elementary School Journal*, *101*, 121–165.

11

Beyond Literal Comprehension: A Strategy to Promote Deep Understanding of Text

Alison King
California State University San Marcos

This chapter presents a comprehension strategy designed to promote deep comprehension of material read. The strategy, "ASK to THINK-TEL WHY©®", is described, its theoretical basis elaborated, and several experimental classroom research studies are reported that demonstrate its effectiveness. The essence of the strategy is a process of reciprocal peer questioning-answering in which the particular questions asked are designed to elicit self-explanations and inferences. When used in peer learning contexts the strategy guides learners to engage in the kind of discussion that supports their construction of representations of text meaning that are coherent at both local and global levels.

When reading text material, the reader creates an understanding of what is being read. This *meaning-making*—this comprehension process—entails the construction of a mental representation of the information in the text, and this representation can in turn be accessed later, when memory for the material is called for. A reader's successful comprehension of text material is evidenced by a representation that, at a minimum, is coherent enough to account for all the information (the main ideas and details) explicitly presented in the text. However, there are different levels of comprehension (and corresponding levels of coherence in mental representations). These levels of comprehension are sometimes referred to as *literal* versus *inferential comprehension*, or *shallow* versus *deep comprehension* (suggesting the metaphor of a lake and simply skimming its surface as opposed to diving down into its depths).

In this chapter, I present a strategy for going beyond the literal level of comprehension to promote a deeper understanding of text through critical reading, thoughtful analysis, and complex inference generation. Thus, this particular comprehension strategy emphasizes thinking critically about text meanings and intentionally making connections between text and relevant knowledge of the world beyond that text.

SHALLOW AND DEEP COMPREHENSION

Whereas literal (shallow) comprehension of a text reflects a minimally coherent mental representation, deep comprehension is indicated by a highly coherent, richly integrated, plausible representation. In terms of Kintsch's (1974, 1988) levels of mental representation (surface code, text based, and situation model), shallow comprehension results from processing text at the two lower levels: surface code and text based (see chap. 1, this volume). This level of processing provides a representation that captures only meaning explicitly stated in the text—the what, who, where, and when of the passage. In contrast, deep comprehension is achieved when the reader goes beyond literal comprehension to use the explicit text and that reader's own prior knowledge to construct such understanding as causes to explain why the events recounted in the text occurred, the probable effects of actions taken, the motives behind people's behavior, and the larger point made by the author of the text (see also chap. 1, this volume). This deeper level of processing results in a representation that is a richer, broader situational model (Kintsch, 1974, 1998) of the text's meanings.

Unfortunately, students rarely gain a deep understanding of the materials they read in their school courses; instead, they settle for shallow knowledge, such as lists of facts, definitions of concepts, and other easily memorized material. Often, students' focus on gaining shallow knowledge from text is reinforced by classroom practices that emphasize lectures; teacher questioning that elicits only short, factual answers; and multiple-choice testing designed to reveal factual knowledge (Davoudi, 2005). To truly understand what they have read, readers need to organize this shallow knowledge and go beyond it. In limiting their understanding to the literal, readers fail to pursue the deep explanations, causes, and implications underlying the knowledge presented, which would enable them to organize the knowledge and to grasp the text's deeper meanings, its message, or point (Davoudi, 2005).

Inferencing

A major difference between shallow and deep level comprehension of text has to do with the inferences and other connections generated by the reader (Cain, Oakhill, Barnes, & Bryant 2001). Both the quantity and quality of inferences and connections made during and after reading affect the reader's comprehension and corresponding representation of the text.

Writers generally leave some material implicit in their text as they assume that the reader will easily figure it out (infer it) from the text. This "figuring out" of implicit information is called *inferencing* and is considered to be a central part of the comprehension process (e.g., Vonk & Noordman, 1990). Even simple text passages require the reader to use prior knowledge to fill in details not included in order to understand the text's meaning (Samuels & Kamil, 1984). By automatically tapping in to their existing knowledge, most readers can make the simple inferences (e.g., bridging inferences) necessary to build a minimally coherent representation of the text (one that captures explicit meaning). Inferencing at this literal level of comprehension occurs automatically, without thinking, and the inferences made are simple, text-based ones. In contrast, deeper level comprehension involves making complex inferences (e.g., causal, elaborative, and predictive inferences). Such inferences go beyond the explicit text to link material within the text to relevant prior knowledge of the world to arrive at the central message of the passage. In fact, several theories of comprehension see this kind of inference generation as the central factor in the construction of situation models of representation (e.g., Graesser, Singer, & Trabasso's [1994] constructionist theory; Kintsch's [1988] construction-integration model; van den Broek, Young, & Linderholm's [1999] landscape model).

Unfortunately, many readers do not automatically make the kind of complex inferences required to build the situational representations typical of deeper levels of comprehension. This kind of inferencing requires effort on the part of the reader, and it is not only effortful, and thus intentional, but it is also often strategic. Lack of competency in the kind of strategies that promote deep comprehension may be one cause of readers' failure to move from shallow to deep comprehension. One such strategy, ASK to THINK–TEL WHY©®,[1] is presented in this chapter (see also chaps. 7, 16, and 14, this volume).

Standards of Coherence

A related reason that many readers do not strive for deep comprehension may have to do with their dispositions toward comprehension and their consequent standards of coherence for their mental representations. Many readers settle for constructing only an understanding that makes sense of the material based

[1]"ASK to THINK–TEL WHY©® is a registered trademark, and the learning procedure itself is copyrighted by Alison King (1994a, 1997, 1998, 1999). Neither the names ASK to THINK–TEL WHY©® or ASK to THINK nor the particular learning procedure known by that name and described herein may be used for any commercial, teaching, or training purpose or any other purpose whatsoever without prior written permission from Alison King.

on what is in the text, that is, an understanding that "hangs together" and accounts for all the text-based main ideas. This literal level of understanding leaves them with a (mistaken) sense that they understand the material they have read (Baker, 1985; Long, Oppy, & Seely, 1994; Otero & Kintsch, 1992), and they feel no need to delve further. In contrast to this makes-sense standard for coherence, other readers strive for full coherence, both for the passage read (local coherence) as well as how it connects to the world outside the text (global coherence). Such readers are driven to generate explanations and other inferences to account for why text-based events, actions, and interactions occur and to explain how they might relate to the situational context external to the text. Such a high standard of coherence exemplifies a disposition toward deep comprehension.

This notion of readers' differing dispositions toward shallow or deep comprehension is consistent with Perkins's conception of a *makes sense epistemology* compared to a *critical epistemology* (Perkins, Allen, & Hafner, 1983). Perkins et al. (1983) made a distinction between learners who adhere to a makes-sense epistemology (whereby the learner tends to minimize cognitive load when building a mental model of a statement by making no further effort to think about it once it appears to make intuitive sense) and those who adhere to a critical epistemology (whereby learners ask when and why their mental model might be subject to criticism; consider alternative hypothetical scenarios, counter arguments, conflicting information, and the like; and as a result build more robust, richly integrated, highly coherent situational mental representations). (See chap. 1, this volume, for more on readers' standards of coherence, and chap. 10, this volume, for a discussion of other motivational aspects of comprehension.)

Strategies for comprehension can be taught and learned (e.g., McGee & Johnson, 2003; Pressley et al., 1992) and, according to Perkins, Jay, and Tishman (1993) and Perkins, Tishman, Ritchart, Donis, and Andrade (2000), positive dispositions toward thinking and learning can be promoted and acquired. It seems reasonable to expect that in the effort to encourage deeper understanding through learning and applying comprehension strategies, readers' resulting strategy proficiency would not only improve their deep comprehension but may also enhance their disposition toward deep understanding and higher standards of coherence.

The strategy presented in this chapter is designed to promote learners' deep comprehension skills as well as positive dispositions toward deep comprehension. The strategy is designed to guide learners in developing richly integrated, situational representations of material read, listened to, or encountered in nontext media, such as TV, movies, advertisements, and computer software programs. The strategy and its theoretical basis are presented. Experimental classroom research studies are reported that demonstrate the strategy's effectiveness.

THE ASK to THINK–TEL WHY©® STRATEGY

ASK to THINK–TEL WHY©® (King, 1994a, 1997, 2007; King, Staffieri, & Adlegais, 1998; King et al., 1996) is a comprehension strategy that involves learners asking thought-provoking questions (the ASK to THINK part) that elicit explanations and inferences (the TEL WHY part) about material to be understood. The ASK to THINK–TEL WHY©® strategy is designed so that the structured discussion resulting from its use in a collaborative learning context is intended to induce in learners a variety of cognitive and metacognitive processes that in turn are expected to enhance comprehension of material discussed. The theoretical basis of this approach is the constructionist theory of comprehension (Graesser et al., 1994). According to the constructionist theory, why-questions and their expected resulting explanations are basic to the construction of meaning (see also chap. 1, this volume).

The ASK to THINK–TEL WHY©® strategy is a more complex version of a previous strategy, Guided Peer Questioning (King, 1989, 1994b), which was found to promote deep comprehension by guiding learners' thoughtful discussion of material (King, 1994b; King & Rosenshine, 1993). Specifically, the Guided Peer Questioning strategy guides learners in how to ask each other thought-provoking questions about what they have read, heard, or seen and provides guidance and support in how to respond by constructing thoughtful, elaborated responses; learners also learn to ask and answer comprehension questions and to sequence their questioning from comprehension to thinking. In the Guided Peer Questioning strategy, learners are provided with a set of general literal comprehension question starters (see Figure 11.1) and a set of open-ended, thought-provoking questions (referred to as *thinking questions* or *connection questions*) such as "How are … and … similar?" and "Explain why … " Individuals select several of these content-free questions to use in guiding them to generate their own content-specific questions on the material being studied (e.g., "How are Shintoism and Buddhism similar?" and "Explain why the storm is relevant to the theme of the power of fear, hate, and prejudice in *Snow Falling on Cedars*"). Then, with a partner or small group, they ask and answer each other's questions.

Learners are also trained in a procedure for generating appropriately elaborated responses (King, 1994b) so that their answers to the thought-provoking questions are more likely to be at a comparably thoughtful level. This component of Guided Peer Questioning is called *TEL WHY* to emphasize the generating

comprehension questions

Describe . . . in your own words.
What does . . . mean?
Why is . . . important?
What caused ... ?

Thinking questions (connection questions)

Explain why . . . .
Explain how . . . .
How are . . . and . . . similar?
What is the difference between . . . and . . .?
How does ... affect ... ?
What are the strengths and weaknesses of ... ?
How could . . . be used to . . . ?
What would happen if . . . ?

Figure 11.1. Generic comprehension and thinking questions used in the Guided Peer Questioning strategy.

explanations aspect of the strategy. Figure 11.2 shows the TEL WHY acronym for this procedure along with reminders for learners regarding what each letter refers to and how this component of the strategy is used. The acronym TEL WHY is intended to emphasize to learners the importance of telling how and why, using their own words to do so, and connecting the idea being explained to something already known. The acronym also keeps learners attention focused on these characteristics of an effective explanation.

Thus, Guided Peer Questioning uses learner-generated questioning to elicit such cognitive activity as self-explanation, inferencing, speculation, elaboration, and making connections between the text and relevant prior knowledge of the world beyond the text. According to both the constructionist (Graesser et al., 1994) and construction-integration (Kintsch, 1988) theories of comprehension, such cognitive activity is essential to meaning-making and promotes building coherent, highly integrated mental representations.

The ASK to THINK–TEL WHY©® strategy uses these same questioning-answering features of Guided Peer Questioning and builds on them by incorporating additional question types, specific question sequencing, and increased emphasis on metacognition. Also, when used in peer learning, two other components are included in the model: (a) supportive interpersonal communication skills and (b) structured reciprocal questioner–explainer roles. ASK to THINK–TEL WHY©®, Guided Peer Questioning, and other versions of the strategy can be used individually by learners; however, they are designed

T	Tell --	what you know to your partner
E	Explain --	the why and the how about something -don't just tell what it is or describe it
L	Link --	connect what you are explaining to something your partner already knows so they'll be sure to understand
W	Tell	Why
H	tell	How
Y	use	Your own words

Notes: adapted from King, 1994b

Figure 11.2. TEL-WHY explanation procedure component of ASK to THINK–TEL WHY©® showing the acronym, its meaning, and how the procedure is used.

primarily to be used in peer learning contexts by same-ability, same-age learners. These strategies have been used successfully in classrooms with students at the university level, in high school, in middle school, and as young as fourth grade. Successful use of the strategy requires extensive training, guidance, and application practice in skills of question asking, question sequencing, explaining and inferencing, and use of supportive communication skills.[2]

Purpose of ASK to THINK–TEL WHY©®

The purpose of this strategy is to promote deep understanding of presented material. In a reading comprehension context, the strategy prompts learners' intentional critical reading; that is, it induces their effortful cognitive and metacognitive processing of text. Research has shown that the cognitive activities that contribute most to deep comprehension include asking appropriate thought-provoking questions (e.g., Graesser & Person, 1994; King, 1989, 1994b; Lepper, Aspinwall, Mumme, & Chabey, 1990), self-explanation (e.g., Chi, deLeeuw, Chiu, & LaVancher, 1994), inferencing (e.g., Graesser et al., 1994), argumentation (e.g., Kuhn, 1991), resolving conceptual discrepancies (e.g., Piaget, 1985), and elaboration in general (Webb, 1989, & Webb & Farivar, 1994). ASK to THINK–TEL WHY©® was designed to induce these

[2]Materials for training and/or use of ASK to THINK–TEL WHY©® are copyrighted and are available only with the written permission of Alison King.

same kinds of cognitive activity in collaborating pairs through a question-asking and -answering process. Learner questioning and responding are guided so as to first activate prior knowledge and consolidate text-based meaning and then elicit beyond-text inferencing, explanation, integration of ideas, and connections among ideas and information within the text and between the text and relevant prior knowledge of the world beyond the text. According to Graesser et al.'s (1994) constructionist theory of comprehension, this kind of mental activity is fundamental to constructing understanding, and according to Kintsch's (1998) construction-integration theory, this cognitive activity results in learners' construction of richly integrated, coherent mental representations of the whole situation (i.e., situation model) implied by the text. Furthermore, comprehension of new information updates mental representations, which in turn have an effect on subsequent comprehension. According to van den Broek et al.'s (1999) *landscape model* of comprehension, this bidirectionality of the relationship between the learner's mental representations and that learner's comprehension process promotes ongoing deep comprehension.

The ASK to THINK–TEL WHY©® strategy is also intended to encourage self-regulation of deep comprehension (e.g., King, 2007). The metacognitive aspects of the strategy—in particular, the skills of monitoring comprehension, reasoning, and learning—are expected to be internalized by readers for later use independently to promote their deep comprehension. Together, then, proficiency with the strategy ASK to THINK–TEL WHY©®, along with the reader's ability to self-regulate use of the strategy, is expected to promote a disposition toward deep comprehension.

Features of ASK to THINK–TEL WHY©®

In essence, learners are taught how to ask questions that elicit explanations and inferences about material to be understood, how to answer questions with relevant thoughtful responses (e.g., explanations and inferences), how to build on each other's responses, and how to assess and monitor each other's understanding. To guide their questioning, learners are provided with generic questions, which they then use as a model to generate their own questions, which are specific to the material read. The ASK to THINK–TEL WHY©® strategy is used primarily after reading (for postreading meaning-making); however, after much experience with the strategy, some readers begin to automatically engage in mental questioning while reading and thus use the strategy for text processing online during reading. When in a peer learning context, learners pose their specific questions to their learning partners, who are guided by the specificity of the questions to generate relevant responses. Questioning is

Ask to

T **THINKING-ABOUT-THINKING QUESTIONS**
(asks learner to "unpack" thinking and self-monitor thinking/ comprehension)

H **HINT QUESTIONS**
(gives clues, reminders & partly framed answers--without disclosing answer—so partner can construct answer)

I **INTELLIGENT THINKING QUESTIONS**
(asks for inference/explanation/connection/application/comparison/insight --induces both partners to engage together in critical thinking & analysis –going beyond facts and concepts by integrating, inferencing, interpreting, explaining, justifying, speculating, comparing)

N **NEED-TO-UNDERSTAND BETTER QUESTIONS**
(probes partner to expand, elaborate, clarify)

K **KNOWLEDGE-REVIEW QUESTIONS**
(asks partners to work together to: 1. activate prior knowledge; 2. retell knowledge e.g., facts, concepts, definitions, processes, formulas, principles, rules; 3. consolidate knowledge base; and 4. monitor comprehension)

THINK ©® adapted from King, 1994a

Notes: The full materials for training and/or use of ASK to THINK –TEL WHY©® are copyrighted and are available only with the author's written permission. ASK to THINK –TEL WHY©® is a registered trademark and the learning procedure itself is copyrighted by Alison King, 1994a, 1997, 1998 and 1999. Neither the names ASK to THINK –TEL WHY©® or ASK to THINK©® nor the particular learning procedure known by that name and described herein may be used for any commercial, teaching, or training purpose or any other purpose whatsoever without prior written permission from Alison King.

Figure 11.3. THINK component of ASK to THINK–TEL WHY©® showing (bottom to top) the function of questions, cognitive processes they are expected to promote, and their intended sequence.

sequenced to build on previous responses and questions so as to move the cognitive activity to progressively higher levels. A metacognitive process for monitoring comprehension and learning is built into the strategy to promote self-regulation of deep comprehension.

Question Asking. Appropriate questioning is at the heart of this strategy. ASK to THINK–TEL WHY©® uses five kinds of questions: (a) review, (b) thinking, (c) probing, (d) hinting, and (e) metacognitive. Learning partners carefully sequence these questions (from knowledge review through thinking to metacognitive questioning) to scaffold their learning from comprehension checking and consolidation of prior knowledge to building new knowledge and monitoring thinking.

Figure 11.3 shows the acronym THINK, which exemplifies the five kinds of questions with a vertical arrow showing the sequence in which questioning proceeds—from bottom to top—from knowledge review, through thinking, to metacognitive questioning. The acronym THINK is designed to portray the process and define what the questions actually do; therefore, the acronym itself prompts users to engage in the appropriate cognitive activity. For example, K stands for *knowledge review*, which entails reviewing relevant prior knowledge; this term reminds questioners to ask review questions and prompts their partners to retell and summarize what they already know about the topic: ideas, definitions, and so on. Probing questions are called *need-to-understand-better questions* (rather than *probing*), as this label captures the intent of those questions. Similarly, thought-provoking questions are called *intelligent thinking* ones, to stress that they call for deep thinking on the part of the one who poses them as well as by the responder. Hint questions are asked throughout, and the term *hint* reminds learners that the question must hint at (but not tell) the answer (see Lepper et al., 1990, for a discussion of the value of asking hint questions). Finally, the expression *thinking-about-thinking* captures what the metacognitive questions ask for (better than would the term *metacognitive*).

As Figure 11.3 indicates, a questioning sequence begins with review questions (e.g., "What causes …?" and "Summarize … in your own words"); in posing these questions, not only are questioners plumbing their partners' memory for the material and assessing their understanding of it, but they are also monitoring their own comprehension. Those review questions activate whatever knowledge partners have on the topic and elicit their retelling of definitions, descriptions, explanations, and elaborations. If an answer to a review question is incomplete, the questioner asks probing questions (e.g., "Tell me more about …") to prompt the explainer to expand on an idea, clarify a point, be more explicit, give an example, or in some other way elaborate. When responses are incorrect or partial, hint questions (e.g., "Have you thought about …?" or "How can … help you?") are asked. Hint questions provide clues or partially framed answers so as to guide explainers to repair any knowledge deficits or errors in reasoning and integrate the modification into their mental representations of the material. The question asked determines the response made, which in turn dictates the next question, both its form (probing, hint, or review) and its content (as the questioner builds on the explainer's response). Thus, at this stage of questioning both partners jointly consolidate the material in memory, monitor their own and each other's comprehension, and repair their knowledge base.

With a shared knowledge base firmly in place, learners proceed to construct new knowledge (e.g., explanations and inferences) onto that base by asking and answering thinking questions (with hint and probing questions as needed). Such questions require going beyond the facts and concepts of the text to induce learning partners to integrate concepts; generate new relationships; make inferences, and explanations, comparison, and insights. Thus, these thinking questions scaffold partners in creating links between ideas within the text and between the text and relevant prior knowledge of the world beyond the text.

Different types of generic thinking questions are used to induce different kinds of cognitive processing, such as explanation (e.g., "Explain how … ," "Explain why …"), inferring cause and effect (e.g., "What do you think causes … ?"), making comparison–contrasts (e.g., "What is the difference between … and … ?") evaluating (e.g., "What is the best … and why?"), using evidence-based reasoning (e.g., "What evidence is there to support … ?"), justifying (e.g., "Do you agree or disagree with this statement: … ? Support your answer"), speculating (e.g., "Explain what disadvantage there might be to using …"), and using logical reasoning ("What might be a counterargument for … ?"). The resulting variety of specific questions generated is expected to prompt different ways of thinking about the material by inducing a variety of cognitive processes in the learner. Thus, asking and answering thinking questions is expected to not only increase the number of connections in learners' knowledge structures but also create a variety of different kinds of connections (e.g., comparative, evaluative, explanatory). Such additional and varied links result in richly connected complex mental representations that are both stable over time and contain numerous and varied cues for retrieval and additional knowledge building (see Kintsch's [1988] construction-integration model of comprehension).

Partners ask metacognitive (thinking-about-thinking) questions throughout the ASK to THINK–TEL WHY©® process. These questions require both questioner and responder to engage in thinking about their own thinking—to monitor the depth of their comprehension. Such questions also help keep partners focused on and mindful of the process. Asking and answering these questions prompts them to "unpack" their thinking. For example, the metacognitive question "What did you learn that you didn't know before?" encourages the partners to self-evaluate their understanding, whereas asking "How did you figure that out?" prompts them to analyze how they figured out an inference or what strategy they used to remember a definition; and asking "How will you remember this?" helps them integrate the material and set up retrieval cues. Such metacognitive questioning provides learners opportunity to become aware of and practiced in monitoring and regulating their own comprehension process.

Question Sequencing. A questioning-answering sequence begins with review questions and proceeds to more sophisticated thought-provoking questions, with hint and probing questions as well as metacognitive questions interjected as needed. The questions, when posed, prompt partners to make corresponding responses. In this way, learners continuously help each other build on their own and each other's previous contributions so as to "scaffold" knowledge construction (Vygotsky, 1978). Question sequencing from review questions through thinking and metacognitive questioning and responding serves to both control learners' progression to deeper levels of comprehension and monitor the depth and extensiveness of that comprehension.

TEL WHY. Learners are also taught the TEL WHY explanation/inferencing procedure that guides them to generate self-explanations, inferences, and other elaborated responses to questions (see Figure 11.2).

Reciprocal Roles. Learning partners exchange roles as needed. When in the teaching role (the ASK to THINK role), a learner is called the *questioner*, and only asks questions (and does not explain or give answers). The learning partner in the TEL WHY role explains (tells why and how), makes inferences, and elaborates (makes connections) and is referred to as the *explainer*.

Supportive Communication. Several communication skills focus on the role of the questioner. These include listening attentively (using eye contact, nodding, saying "Uh-huh"), providing thinking time by waiting in silence after asking a question before expecting a response (Graesser & Person, 1994; Rowe, 1986), giving feedback on the accuracy and completeness of an answer, and giving encouragement as needed.

RESEARCH ON ASK TO THINK–TEL WHY©®

The results of a program of research on the effectiveness of Guided Peer Questioning with provided thought-provoking generic questions revealed that the procedures can promote deep comprehension in students from fourth grade to the university level (King, 1989, 1994b; King & Rosenshine, 1993; King et al., 1998; King et al., 1996). In the following sections, I describe four experimental design classroom studies that trace the evolution of the strategy and exemplify the effectiveness of different components of the strategy. The first three studies are summarized, and the third, which addresses the role of the metacognitive-questions component, is more fully elaborated.

Effects of Elaborated Thought-Provoking Questions

In an early study to assess the effectiveness of the thought-provoking questions component of the Guided Peer Questioning (King & Rosenshine, 1993), the elaborated version of the questions was compared to a less elaborated version (why- and how-questions) and an unguided questioning condition. Partners in one guided questioning conditions were trained to generate thought-provoking questions using very structured generic questions such as "Why is … important?" and "What would happen if … ?" (the highly elaborated question stems), and partners in a second condition generated thought-provoking questions using question starters such as "Why … ?" and "How … ?" (the less elaborated question stems) to guide their question generation. Partners in an unguided questioning (control group) condition were simply directed to ask (and answer each other's) questions without guidance.

Deep comprehension in this study (and in subsequent studies) was measured directly in two ways: (a) by the inferences students made during pair discussion sessions (number of inferences generated per minute of discussion) and (b) by students' scores on the inference items on written subject-matter tests. An *inference* in this study was defined broadly as a statement that integrated aspects of the new material in some manner or in some other way went beyond the material presented to show complex constructed meaning (meanings not explicitly stated in the text and requiring inferencing/explanation based on material that was explicitly provided). Therefore, explanations, inferences, interpretations, relationships between ideas, justifications, speculations, and statements linking the text content to prior knowledge were coded as inferences. For example, during discussion a fourth-grader made this analogy for the nervous system: "It is like the school office phone. The office phone has different lines—there are a lot—they can all be used at once without interfering with each other—to send messages back and forth" (from King, 1994b); this statement was coded as inferencing/explanation because it explains a concept by linking the new information about the nervous system to the student's prior knowledge about telephone systems.

Fourth-graders who used highly elaborated question stems outperformed those who used less elaborated stems and the unguided questioners on explanations provided during discussion sessions, posttest inferential comprehension, and knowledge mapping. These findings suggested that, in peer discussion contexts, structured guidance in asking thought-provoking questions elicits explanations and inferences, indicators of deep comprehension.

Effects of Thought-Provoking Questions of Differing Focus

Another study (King, 1994b) compared the effects of two sets of thought-provoking questions with differing focuses: (a) internal to the text passage (text based) versus (b) both internal and external to it (text- and experience-based questions). Two versions of the Guided Peer Questioning strategy and unguided questioning were compared in a peer learning context. Forty-eight fourth- and fifth-graders studying science material in one large combined class were randomly assigned to pairs, which were then randomly assigned to three peer-questioning conditions. Students worked in pairs to understand the material through discussion by asking and answering each other's self-generated questions. In one guided questioning condition (text-based questioning), students' discussions were guided by the use of generic questions: knowledge questions (literal comprehension questions) as well as thought-provoking questions (e.g., "How are … and … similar?" and "How does … affect … ?"). The latter questions were designed to promote explanation and inferencing within the presented material, as they asked partners to make connections among ideas within the text; thus, they were expected to promote building of text-based representations. In the second condition (experience-based questioning), students' discussions were also guided by the use of generic knowledge questions and thought-provoking questions; however, some of those thought-provoking questions were text based, and some were experience based. Experience-based questions (e.g., "How does … tie in with … that we learned before?" and "How could … be used to … ?") were expected to prompt learners to deliberately access their relevant prior knowledge and link the new material to their existing knowledge structures and thus more likely promote the building of situation models. In a control group (unguided questioning), students were not trained in questioning but directed to engage in questioning. Students in all three conditions were trained in the TEL WHY procedure to support them in generating explanations and inferences. Thus students were trained to ask for and provide explanations and inferences about the material being studied.

Deep comprehension in this study was measured as in the previous study by King and Rosenshine (1993); that is, by the inferences students made during pair discussion sessions and by students' scores on the inference items on written subject-matter tests. Posttest results in this study indicated that students using the experience-based questioning strategy generated significantly more inferences on written tests as well as during their pair discussion itself[3] than did the text-based questioners, who in turn significantly outperformed the unguided questioners. Furthermore, on a delayed written retention test, experience-based questioners retained the inferences they had generated significantly better that did text-based questioners who, in turn, outperformed control participants. Knowledge mapping by students was used to capture their mental representations of the material. The knowledge

[3]Verbal interaction during peer strategy sessions was taped, transcribed, and analyzed in all studies.

maps of the two trained questioning conditions were more coherent and richer than those from the unguided questioning condition (e.g., more connections, more variety in labels of links), and experience-based questioners' knowledge maps were significantly more representative of the situation implied by the material than were those of text-based questioners.

This study indicates that complex inferencing can be promoted by the guided-knowledge-and-thinking question component of this particular questioning-and-answering strategy. The results further show that when the questioning component of the strategy focuses on experience-based questions in particular, it can better support learners in building representations that capture the situation implied by the text.

Role of Question Sequencing

An extension of that basic guided questioning strategy was developed (King, 1996) and called *ASK Your Partner to Think*. That strategy includes four components. In addition to the questioning component (using knowledge review and thinking questions), and the explaining component (TEL WHY), two new kinds of questions (probing and hinting) were added to the questioning component, and two new components were added to the strategy: (a) supportive communication skills and (b) sequencing of the (now) four kinds of questions. The hint and probing questions and the sequencing component were added to the strategy based on findings from studies of expert tutoring (e.g., Lepper et al., 1990) showing that asking probing and hint questions facilitates elaboration and that sequencing questions helps scaffold learning (see also Vygotsky, 1978).

To determine the effectiveness of components of the ASK Your Partner to Think strategy, King et al. (1998) compared seventh-graders in three conditions working in pairs to understand science material. All students (n = 58) were trained in the TEL WHY explanation/inferencing procedure and in use of the supportive communication skills. Students were randomly assigned to pairs and then to conditions. Control group pairs were trained in use of the explanation/inferencing skills and communication skills during peer discussion. In addition to these components of the strategy, students in a second condition were trained in asking knowledge review and thought-provoking questions on the material. Students in the third condition (sequenced questioning) received additional training in how to ask four kinds of questions (knowledge review, hint, probing, and thought-provoking questions) and in how to sequence their questioning effectively. Thus, the strategy used in the third condition was composed of four components: (a) basic knowledge review and thinking questioning; (b) TEL WHY; (c) supportive communication skills;

and (d) sequencing of questions, including use of probing and hinting questions. In contrast, the strategy used in the first condition consisted of two components (TEL WHY and supportive communication skills), and that of the second condition contained three components (TEL WHY, supportive communication, and knowledge review and thought-provoking questions).

On written tests of deep comprehension for subject matter at posttreatment, and on a transfer test, as well as on an 8-week follow-up retention test, students in the sequenced-questioning condition generated significantly more complex inferences than those in the other two conditions. Sequenced-questioning students also outperformed the other groups on generating inferences during their tape-recorded pair discussions at posttest and transfer. These findings suggest that the sequenced-questioning-with-four-kinds-of-questions component of this guided questioning strategy adds a measure of effectiveness for deep comprehension over and beyond that of the basic Guided Peer Questioning strategy.

A Role for Metacognitive Questioning

The ASK Your Partner to Think version of the strategy had even more metacognition built into it than the Guided Peer Questioning version because of ongoing comprehension support of the probing and hint questions as well as additional comprehension checking by both partners during question sequencing, and thus it allowed for more joint and individual monitoring of comprehension. However, it was presumed that if partners were to ask metacognitive questions per se, then this might improve monitoring of their comprehension and promote deeper comprehension, in particular, complex inferencing.

Accordingly, the next incarnation of the strategy (now called ASK to THINK–TEL WHY©® involved adding a fifth component: explicit metacognitive (thinking-about-thinking) questioning. Metacognitive questions (e.g., “How did you arrive at that answer?” and “What made you think about that?”) coming on the heels of a partner’s inferences presumably might enhance both partners’ awareness of their comprehension processes (i.e., it would make the process obvious to learners) and lead them to monitor and regulate the current state of their comprehension more carefully. Such an explicit emphasis on self-regulation of their own comprehension process might in turn further improve learners’ understanding by promoting deliberate inferencing as well as ongoing evaluation of representations they were constructing. Several studies have shown that students who were trained to formulate metacognitive questions for one another during group work improved their problem-solving success (e.g., King, 1991; Mevarech & Kramarski, 1998) and comprehension (e.g., Dominowski, 1998; Paris & Winograd, 1990; also see Hacker, 1998, for a summary of research on effects of metacognition on learning). With the addition of the metacognitive-questioning component, as well as asking and answering thought-provoking questions, students would need to reflect on their responses to those questions by answering a follow-up metacognitive question; that is, in addition to self-monitoring of comprehension, they were called on to consciously and deliberately think about their thinking: how they figured out their answers, how they might remember what they had found out, and where they might use it again.

A study was conducted to assess the effectiveness of this added metacognitive component of the strategy. The study took place in a typical classroom context over an 8-week period, ending in mid-June, when students went on summer break. A 3-month follow-up testing session was conducted when students returned to school in September. Twenty-six fourth-graders studying social studies were randomly assigned to same-gender pairs, and pairs were then randomly assigned to two conditions: sequenced-questioning with metacognitive questions (using ASK to THINK–TEL WHY©®) and sequenced questioning (using ASK Your Partner to Think—which is exactly the same as ASK to THINK–TEL WHY©® but without the metacognitive-questioning component). Students in both conditions received the same training and practice in the four strategy components: (a) generating and asking four kinds of questions (knowledge review, thinking, probing, and hint questions); (b) sequencing those questions from review to thinking with need-to-understand-better (probing), and hinting questions interspersed as needed; (c) using the TEL WHY explanation/inferencing procedure; and (d) using supportive communication skills. Students in the sequenced-questioning-with-metacognitive-questions condition received additional training and practice in asking a fifth kind of question: metacognitive questions.

As in the previous studies cited, deep comprehension was measured by students' scores on the inference items on written subject-matter tests and by the rate at which they made inferences during actual pair discussion sessions. Metacognitive questioners significantly outperformed the other students on inferencing during discussions in both the posttreatment session as well as at transfer (when they no longer had reminders to use their questioning strategies). On written tests of inferencing, students in the sequenced-questioning-with-metacognitive-questions condition significantly outperformed those in the sequenced-questioning condition at posttreatment and in a transfer context. Furthermore, at follow-up testing, 3 months after the posttreatment session, the metacognitive questioners were able to make new inferences based on what they could remember from that posttest session material to a significantly greater extent than did students in the other condition (even though there was no significant difference between treatment conditions on retention of literal material from the posttreatment session).

Students' metacognitive awareness of their comprehension processes involved in making inferences on the written tests also was assessed. To determine the extent to which students showed awareness of the thinking they had used to construct their understanding when answering the open-ended inference questions on the written tests, metacognitive questions were attached to two of the inference questions on each test (posttreatment, transfer, and 3-month follow-up tests). For example, after the inference question "In what ways do you think the squatters were like the vigilantes and the miners?", space was left for an answer and, after that space, a metacognitive question—"How did you figure that out?"—was posed as a second part of the question. At both transfer and the 3-month follow-up, metacognitive questioners were significantly more aware of their thinking processes involved in constructing inferences on written tests than were the questioners who were not trained to use metacognitive questions.

The follow-up written comprehension test was based on the posttest session content of 3 months earlier. This test was intended to assess the extent to which students could make new inferences based on what they remembered from that test and to assess their awareness of their thinking while doing so. Although both strategy groups had retained the same amount of material from the posttreatment session (as shown by their scores on the literal comprehension test), the metacognitive questioners scores were much higher than the sequenced-inquiry students on new inferences generated from that material (the inferential comprehension test); their scores on metacognitive awareness also were significantly higher. These results suggest the possibility of a long-lasting (3 months) effect for ASK to THINK–TEL WHY©® when used as a strategy for deep comprehension.

In this study, training and practice in metacognitive questioning presumably improved students' skill in comprehension as well as awareness of their comprehension processes and their regulation of those processes, which in turn enhanced their ability to construct inferences both during their study discussions per se and subsequently on written comprehension tests.

SELF-REGULATED READING COMPREHENSION WITH ASK to THINK–TEL WHY©®

The metacognitive nature of the strategy in general, together with the metacognitive-questioning component, promotes comprehension monitoring and self-regulation of comprehension. Thus, as a peer learning strategy it supports learning pairs in self-regulating their joint comprehension. Although ASK to THINK–TEL WHY©® was designed for peer learning, it is expected that, over time, the strategy will eventually be internalized by individual readers and used independently by them so that they begin to pose and answer their own questions, thus promoting their own reading comprehension in a truly self-regulated manner.

Certain aspects of the strategy are particularly amenable to internalization (and thus self-regulation). For example, with practice, the generic guiding questions could become internalized as Vygotskyian (1978) "inner speech," thus allowing learners to engage in self-talk (e.g., posing the questions to themselves) to prompt their own comprehension processes (Rogoff, 1990; Vygotsky, 1978). Although probing and hint questions would be unlikely candidates for use by readers working alone, self-questioning with internalized thought-provoking and metacognitive questions (and sequencing of those questions) would serve the reader well as skills for digging below the surface of text to achieve deeper comprehension. Because the roles of questioner and explainer are alternated during peer learning with ASK to THINK–TEL WHY©®, practiced learners would gain experience in both posing questions and generating inferences/explanations, so that over time both roles could be internalized. Peer modeling of each role would help to make the roles easier to remember and assume later. Internalizing the questioning and sequencing procedures could support learners in becoming self-regulated in their reading comprehension; in essence, they could prompt their own execution of the entire questioning-answering strategy as they read independently.

When readers achieve proficiency with ASK to THINK–TEL WHY©® (whether as a pair or alone), and are able to self-regulate use of the strategy through internalization (or as a pair), it should be easier for them to move from shallow to deep comprehension. Once facility with the strategy is attained and the ability to self-regulate it becomes automatic, readers' disposition toward deep comprehension would presumably be enhanced.

CONCLUSION

ASK to THINK–TEL WHY©® is a strategy designed to support deep comprehension of material read, heard, and seen. In this chapter, I have attempted to show how that strategy functions to guide readers in constructing representations of text meaning that are coherent at both local and global levels. The strategy achieves this end in general by inducing a variety of cognitive and metacognitive processes in learners during and after reading from text. More specifically, the major thrust of the strategy is its emphasis on eliciting learner explanations and inferences as a result of their asking thought-provoking questions relevant to the text being processed while continuously monitoring their comprehension. In essence, learners are taught to ask five different kinds of questions and to sequence those questions according to their partners' response and comprehension level in order to build on each others questions and responses. Doing so is expected to guide the ensuing discussion so that it moves to progressively more complex levels of thinking and metacognition—from review of previously learned material to comprehension of new material to generating explanations and inferences to monitoring thinking and comprehension. In this way, learners use the strategy to actually scaffold each other's learning. This strategy is compatible with Graesser et al.'s (1994) constructivist theory of text comprehension because of its emphasis on asking questions that elicit explanations and inferences as a means of constructing meaning and building coherent representations.

Although the theoretical basis of ASK to THINK–TEL WHY©® is the constructivist theory of text comprehension, findings from studies comparing the individual components of the strategy may also suggest some possible elaboration of that theory. The ASK to THINK–TEL WHY©® strategy evolved and became refined over the span of several research studies that evaluated the effectiveness of the strategy and its individual components. Soon after the initial version of the strategy, Guided Peer Questioning, was developed and assessed in classroom learning contexts (King, 1989, 1990), the notion arose that the strategy might be just as effective if simple "why?" and "how?" questions were used to prompt learner questioning rather than the elaborated thought-provoking questions; that is, those simple questions might be as thought-provoking and might just as readily guide learners in generating their own specific questions on material read. In a study with fourth-graders comparing the effectiveness of these two kinds of questions, King and Rosenshine (1993) found that the more elaborated questions were indeed more thought provoking—presumably because they provided more specific guidance to learners in generating questions highly relevant to their readings and thus resulted in more explanation and inferencing during discussion and on written tests. With the effects of elaborated thought-provoking questions established as more powerful for building knowledge (explanation and inferences) than the less elaborated why-and how-questions, the research agenda moved forward to examine how effects of thought-provoking questions focused on text only might compare to similar questions with an external focus emphasizing learners' prior knowledge. Would learners using text-based questions in the strategy construct situation models that were equivalent to those of students using a combined set of text-based and experience-based questions? The results of that study (King, 1994b) showed that fourth- and fifth-graders using the experience-based questioning strategy generated significantly more explanations and inferences during peer interaction and on written tests than did the text-based questioners or controls. Knowledge maps generated by the experience-based questioners indicated that their mental representations of the material were more representative of the situation implied by the material than were those of text-based questioners. Results of these two studies support the "explanation assumption" of Graesser's constructionist theory of text comprehension (Graesser et al., 1994; chap. 1, this volume) in that they substantiate other studies that have found the value of using questioning to elicit explanation as a means of building meaning from text. However, it should be noted that the constructionist theory might be warranted in placing more emphasis on the kinds of why-questions learners ask in their quest for deep comprehension, as elaborated specific ones appear to be very effective.

Two other aspects of the strategy that were validated as ASK to THINK–TEL WHY©® evolved are the sequencing of questions and the use of explicit metacognitive questions. The sequenced-questioning-with-four-kinds-of-questions component was added to the strategy to more carefully control how knowledge questions as well as probing and hint questions could be used by learners to establish a shared knowledge base prior to building new knowledge on that base through use of the thought-provoking questioning. When seventh-graders used the refined strategy to carefully and consistently sequence their questioning-answering discussion, their resulting deep comprehension was superior to that of their grade mates who used the basic Guided Peer Questioning strategy without sequencing of the four kinds of questions (King et al., 1998). The final component added to the strategy (now ASK to THINK–TEL WHY©®) was metacognitive questioning. The study in which the use of ASK to THINK–TEL WHY©® with the metacognitive questions component was compared to the strategy without that component showed that the role of explicit metacognition in the strategy is important to meaning making. Perhaps as research progresses on this front, Graesser might find that adding a "metacognitive assumption" to the constructivist theory would be appropriate.

In this chapter, I have presented the ASK to THINK–TEL WHY©® strategy and discussed it in the context of reading and text comprehension. However, it also has potential as a strategy for comprehending material heard (e.g., in lectures, demonstrations, and teacher lessons) and material viewed onscreen (e.g., movies, television programs, PowerPoint presentations, advertisements, and computer software programs).

Although research studies using ASK to THINK–TEL WHY©® have been conducted only with fourth-graders and seventh-graders, the strategy is assumed to be appropriate for use with any age and grade level above fourth grade. The subject areas used in research studies were social studies and science; however, it is presumed that the strategy would be equally successful with any expository or literary material. Future studies are planned to assess its effectiveness with students reading literature.

REFERENCES

Baker, L. (1985). Differences in standards used by college students to evaluate their comprehension of expository prose. *Reading Research Quarterly*, *20*, 298–313.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, *29*, 850–859.

Chi, M. T. H., deLeeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self explanations improves understanding. *Cognitive Science*, *18*, 439–477.

Davoudi, M. (2005). Inference generation skill and text comprehension. *The Reading Matrix*, *5*, 106–122.

Dominowski, R. L. (1998). Verbalization and problem solving. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 1–23). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, *31*, 104–137.

Graesser, A., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *35*, 371–395.

Hacker, D. (1998). Definitions and empirical foundations. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.) *Metacognition in educational theory and practice* (pp. 1–23). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

King, A. (1989). Effects of self-questioning training on college students' comprehension of lectures. *Contemporary Educational Psychology*, *14*, 1–16.

King, A. (1991). Effects of training in strategic questioning on children's problem-solving success. *Journal of Educational Psychology, 83*(3), 307-317.

King, A. (1994a). ASK to THINK–TEL WHY©®. Unpublished manuscript and materials, California State University San Marcos.

King, A. (1994b). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal*, *30*, 338–368.

King, A. (1996). *A sequenced inquiry model of peer tutoring.* Unpublished manuscript, California State University, San Marcos.

King, A. (1997). ASK to THINK–TEL WHY©®: A model of transactive peer tutoring for scaffolding higher-level complex learning. *Educational Psychologist*, *32*, 221–235.

King, A. (1998). Transactive peer tutoring: Distributing cognition and metacognition. *Educational Psychology Review*, *10*, 57–74.

King, A. (1999). Discourse patterns for mediating peer learning. In A. O'Donnell & A. King (Eds.), Cognitive perspectives on peer learning (pp. 3–37). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

King, A. (2007). Scripting collaborative learning processes: A cognitive perspective. In F. Fischer, H. Mandl, J. Haake, & I. Kollar (Eds.), Scripting computer-supported collaborative learning: Cognitive, computational, and educational perspectives (pp. 13–37). New York: Springer.

King, A., Burton, S., Galloway, L., & Verdusco, E. (1996). *Effects of metacognitive questioning during sequenced inquiry peer tutoring on fourth graders' higher-level learning*. Unpublished manuscript, California State University San Marcos.

King, A., & Rosenshine, B. (1993). Effects of guided cooperative questioning on children's knowledge construction. *Journal of Experimental Education*, *61*, 127–148.

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology*, *90*, 134–152.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A constructive-integration model. *Psychological Review*, *95*, 163–182.

Kuhn, D. (1991). *The skills of argument*. New York: Cambridge University Press.

Lepper, M. R., Aspinwall, L. G., Mumme, D. L., & Chabey, R. W. (1990). Self-perception and social perception processes in tutoring: Subtle social control strategies of expert tutors. In *Self-inference processes: The Ontario Symposium* (pp. 217–237). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1456–1470.

McGee, A., & Johnson, H. (2003). The effects of inference training on skilled and less killed comprehenders. *Educational Psychology*, *23*, 49–59.

Mevarech, Z. R., & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. A*merican Educational Research Journal*, *34*, 365–394.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in text: What readers believe versus what they read. *Psychological Science*, *3*, 229–235.

Paris, S. G., & Winograd, P. (1990). Promoting metacognition and motivation of exceptional children. *Remedial and Special Education*, *11*, 7–15.

Perkins, D. N., Allen, R., & Hafner, J. (1983). Difficulties in everyday reasoning. In W. Maxwell (Ed.), *Thinking: The expanding frontier* (pp. 177–189). Philadelphia: Franklin Institute.

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill–Palmer Quarterly*, *39*, 1–21.

Perkins, D. N., Tishman, S., Ritchart, R., Donis, K., & Andrade, A. (2000). Intelligence in the wild: A dispositional view of intellectual traits. *Educational Psychology Review*, *12*, 269–293.

Piaget, J. (1985). *The equilibrium of cognitive structures: The central problem of intellectual development*. Chicago: University of Chicago Press.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist*, *27*, 91–109.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.

Rowe, M. B. (1986). Wait time: Slowing down may be a way to speeding up! *Journal of Teacher Education*, *37*, 43–50.

Samuels, S. J., & Kamil, M. L. (1984). Models of the reading process. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 185–224). New York: Longman.

van den Broek, P., Young, M., & Linderholm, Y. T. T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. In S. R. Goldman & H. V. Oostendorp (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Vonk, W., & Noordman, L. G. M. (1990). On the control of inferences in text understanding. In D. A. Balota, G. B. Flores D'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 447–464). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, *13*, 21–39.

Webb, N., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal*, *31*, 369–395.

IV

Automated Interventions to Improve Reading Comprehension Strategies

12

Web-Based Reading Comprehension Instruction: Three Studies of 3D-Readers

Mina C. Johnson-Glenberg

The NeuronFarm, Madison, Wisconsin, and Waisman Center, University of Wisconsin—Madison

This chapter describes three studies on a Web-based application for comprehension instruction called 3D-Readers. This application is designed to both instruct and assess young adolescent readers' use of verbal and visual metacognitive strategies and their comprehension of hybrid-style science texts. The five significant findings from the three studies were the following: Study 1, with poor comprehenders, revealed significant gains on constructing answers to open-ended questions over eight sessions. In addition, these readers significantly altered their reading processes when encouraged to reread texts, with the poorer comprehenders rereading more often. Study 2, with students with attention deficit disorder/attention deficit hyperactivity disorder, revealed significant vocabulary gains, and significant gains in self-reported metacognitive strategy use after six sessions. Study 3, with summer school students, also revealed significant gains in vocabulary skill, as well as significant gains in the quality of questions generated over four sessions. The results support the conclusion that question generation and the construction of visual simulations are important strategies for increasing comprehension monitoring and learning.

Decades of research support the finding that strong text comprehenders use strategies when they read (National Reading Panel [NRP], 2000; Palincsar & Brown, 1984; Pressley, 2000). Strategy instruction traditionally has not received much time in the classroom (Durkin, 1978–1979). This situation has not changed appreciably over time. In fact, observational research suggests that strategy instruction is largely absent in classroom-based curricula and pedagogy. Furthermore, when strategy instruction is part of classroom curricula, its quality is often problematic in that the strategies are enacted in a routine fashion and rarely for the purpose of enhancing cognitive engagement (Garcia, Pearson, & Taylor, 2004).

There may be several reasons for this lack of commitment. Strategy training is effortful for the teacher. It can take large chunks of time to properly teach strategies—to model, to assess, to practice the strategies. In addition, teachers need to be specifically trained in the strategies themselves. Strategy use is an internal construct; in order for teachers to teach the strategies to their students and then assess students' mastery, teachers must ask questions of the students and wait for answers. This sort of individualistic cognitive assessment takes time and seems prohibitive to perform with an entire class learning multiple strategies. Teachers must also be explicit and use direct instruction as to when and how to use strategies (Duke & Pearson, 2002); this is especially important for striving readers. Another factor that might explain the dearth of strategy instruction is that it is effortful for students as well. To stop during text comprehension and consciously work on a strategy may seem unnatural at first. However, as the strategy becomes more automatized it should feel more natural. Strategies that are frequently used may become automatic (Elshout-Mohr & van Daalen-Kapteijns, 2002). Striving, and even on-level readers, need reader-friendly texts and stress-free time with which to practice strategies until mastery. Because students assimilate strategies at different rates, quality strategy instruction may not actually be conducive to a whole-class instructional paradigm.

Furthermore, for the students struggling with comprehension, teachers are faced with the task of choosing the optimal strategies for each student's profile. Many variables affect the choice and efficacy of a strategy: type of text, prior knowledge, preferred learning style, amount of repetition needed for mastery, and so on. The task before a classroom teacher with 30 readers, all possessing unique cognitive profiles and varying levels of motivation, can be daunting. The good news is that computers are particularly well-suited to the tasks of individualized instruction, repetitive practice, and immediate feedback. In addition, recent developments in graphic optimization algorithms and increased bandwidths in schools make highly visual, Web-based strategy instruction particularly appealing and easy to implement. *3D-Readers*™ is a grant-funded, Web-based application specifically designed to fill the need for reading comprehension strategy instruction.

WHAT ARE 3D-READERS' STRATEGIES? WHY WERE THEY CHOSEN?

3D-Readers integrates optimal practices in reading instruction and embeds them in high-interest, standards-based content area texts. In the late 1990s, the National Institutes of Health convened a panel of reading researchers, and the resulting meta-analysis from the NRP (2000) lists 16 categories of comprehension instruction. The NRP report (2000) states that 7 of these categories demonstrate a firm scientific basis of efficacy in improving typical readers' comprehension. Our instructional program uses 5 of these 7 strategies: (a) comprehension monitoring, (b) question answering, (c) question generation, (d) summarization (evident in the final constructed answers), and (e) graphic organizers. Special emphasis is given to these first four components. The two strategies not included are semantic organization (other software does this well) and cooperative learning. The decision was made to create a first iteration that did not include collaborative or cooperative learning so that teachers could gather unbiased estimates of individual performance. In addition, mental imagery is positively reviewed in the NRP report, and it has been included in the *3D-Readers* application (see also chap. 9, this volume).

3D-Readers is predicated on the hypothesis that generativity and construction increase learning. *Generative activities* are ones that the user must initiate, evaluate, plan, and execute. Slamecka and Graf (1978) demonstrated the power of the generation effect by showing significantly better learning for students who generated words (associates, synonyms, etc.) versus merely reading the words. Generativity requires an active process (Naps et al., 2003). In the reading domain, the generative model is supported by decades of research on its superiority for training reading comprehension (Wittrock, 1991; Wittrock & Kelly, 1984). Doctorow, Wittrock, and Marks (1978) found that students who generated their own summaries comprehended and retained significantly more than those who did not. Generativity leads students to become more "active and responsible for constructing meaning … by building relationships 1) across subject matter concepts and 2) between subject matter and students' knowledge" (Wittrock, 1991, p. 178). The emphasis on active generation is most resonant with the constructionist framework of text comprehension (chap. 1, this volume; Graesser, Singer, & Trabasso, 1994). Readers comprehend because they partake in three principles: (a) They set and attend to goals; (b) they maintain coherence as they read; and (c) they generate explanations of why things tend to occur—events, actions, author's tone, and so on. One of the best methods with which to train poor comprehenders to monitor their comprehension, to maintain coherence, and to initiate repair strategies is question generation (Rosenshine, Meister, & Chapman, 1996), which is a verbal strategy. In addition, Sadoski and Paivio (2001) and Mayer (2001; Mayer & Sims, 1994)—who work in Web-based multimodal learning—would suggest that strategies should not only reside in the verbal domain. Training and allowing readers to visualize content is another goal of *3D-Readers*. The verbal and visual strategies embedded during the *3D-Readers*' learning experience were chosen because they were both generative and research based. In the following sections, I discuss research related to the application's primary components.

Learning Goals and Prior Knowledge

Every session on *3D-Readers* begins with learning goals so the students are aware of knowledge expectations. They are then prompted with an open-ended prior-knowledge question, constructed to be engaging and age appropriate. Prior knowledge is not only predictive of learning (Recht & Leslie, 1988) but also often critical in helping readers make necessary inferences in texts that might contain gaps (McNamara, Kintsch, Songer, & Kintsch, 1996). Activating prior knowledge before reading can stimulate previous topic schemas and more readily provide a structure for the newly acquired knowledge to adhere.

Vocabulary: Prereading and Postreading Assessment

A report by the RAND group (Snow, 2002) noted that vocabulary instruction is one of the primary components of reading instruction. *3D-Readers* is built on the premise that words need to be encountered in authentic contexts. Thus, in a *3D-Readers*' text, the meaning of a vocabulary word is embedded in the immediate context in which the word is first encountered. Students first perform a 7-item multiple-choice pretest with no feedback. As students read through the text, they encounter all the vocabulary words in a blue-colored font. When they click on the word, they see the definition of the word highlighted in blue in the surrounding text. The definitions are often in the next clause and never more than two sentences away (see Figure 12.1 for an example). After reading the approximately 2,000-word text, students take a vocabulary posttest. In the posttest they receive immediate, item-specific feedback.

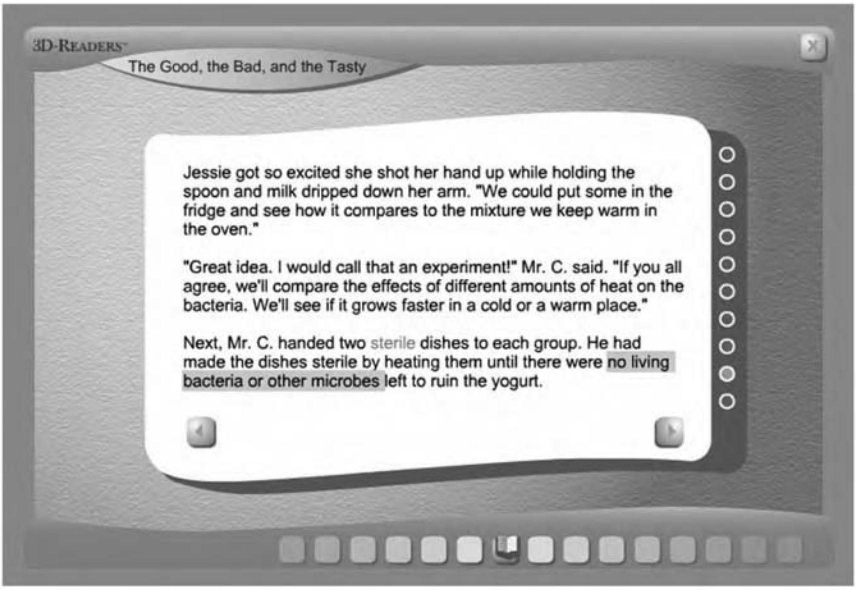


Figure 12.1. Screen shot of the vocabulary word sterile with the highlighted definition.

THE EMBEDDED STRATEGIES: VERBAL AND VISUAL

3D-Readers instructs students in several verbal strategies. The primary embedded verbal metacognitive strategy is *question generation*. This verbal strategy features prominently in Reciprocal Teaching (RT; Palincsar & Brown, 1984), a package of four strategies: (a) clarification, (b) prediction, (c) summarization, and (d) question generation. Rosenshine et al. (1996) conducted a meta-analysis of question generation. They found that question generation training yielded an impressive median posttraining effect size of 0.86 on experimenter-designed measures. Other elements of RT are also found in *3D-Readers*, in that clarification is addressed via vocabulary definition and summarization skills are required for answering the final open-ended questions.

More on Question Generation

Question generation is a powerful tool with which to monitor comprehension. One of the hallmarks of struggling comprehenders is that they do not know that they do not know. Prompting students to stop and create a question, and then to answer that question, trains students in ongoing monitoring. The hypothesis is that, with practice, this strategy of generation and self-assessment will become automatic and internalized.

Before generating questions, students go through a training module that instructs them on the importance of strategies and the elements of a good question. The students take the creation of the questions seriously because they know they must answer one of their questions at the end of the module. Figure 12.2 shows an example of the screen a student would see, with the automatic score on the right-hand side. The student has posed a yes-or-no question. The quality of the two self-generated questions is assessed with the automatic scoring algorithm called *Hi-dimensional Expert Match Algorithm* (HEMA). These scores range from 0 to 6 and are returned to the students in less than 2 seconds. To do this, HEMA first creates large word co-occurrence matrices gathered from the Web. It then plots relevant similarities between high- and low-scoring answers in a K-dimensional space. These similarities are used in a regression equation, and when the fit is poor neural network algorithms are also used to fine-tune the final scores (Johnson-Glenberg, 2005).

Visualization Strategies

Many researchers advocate a multicomponential approach to comprehension instruction (NRP, 2000; Pressley, 2000). Thus, in addition to the embedded verbal strategy of question generation, there is an embedded visual strategy based on generative learning. The ultimate act of reading is the creation of a mental model (Johnson-Laird, 1983) or a situation model (Kintsch, 1988). Although not all mental models are visual, research on the importance of imagery and visualization during information processing and reading has a long and respected history. Based primarily on Paivio's *dual code theory* (Paivio, 1986; Sadoski & Paivio, 2001), numerous studies report significant comprehension gains by experimental groups that were either encouraged or taught to visualize while reading (Bell, 1991; Gambrell & Jaywitz, 1993; chap. 9, this volume; Johnson-Glenberg, 2000; Levin, 1973; Mayer & Sims, 1994; Oakhill & Patel, 1991; Pressley, 1977). The hypothesis is that constructing a visual model on the screen should aid striving readers in building internal, visual models. By repeatedly manipulating and building images, readers discover and/or confirm how the segments of sequential text fit together in a three-dimensional gestalt (hence the name *3D-Readers*).

Merely viewing an animation does not automatically lead to better comprehension (Hegarty, Narayanan, & Freitas, 2002). As such, many results from visualization studies in technology can appear ambiguous (Naps et al., 2003). Splitting studies into categories of low and high generativity clarifies the results. Of 21 experiments that Naps et al. (2003) reviewed in a meta-analysis,

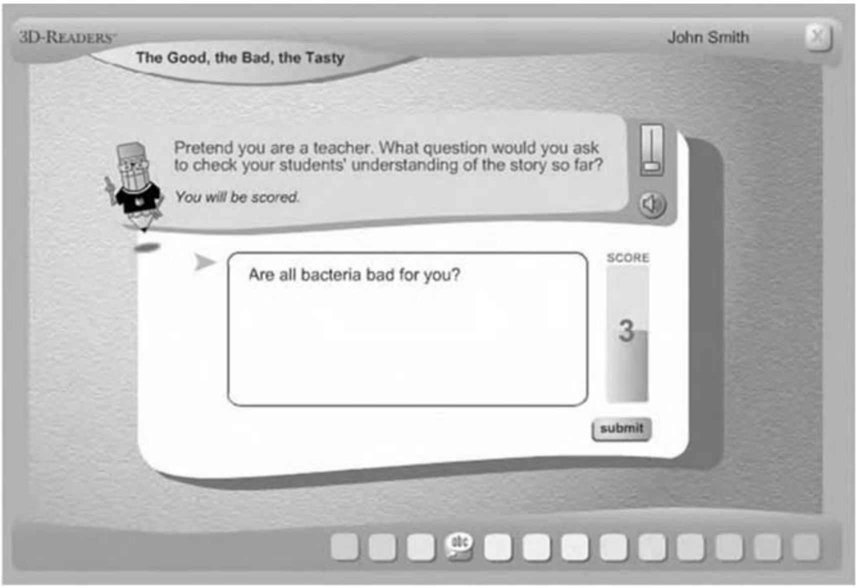


Figure 12.2. Sample question generation screen shot.

only 12 used "learner involvement," 10 of which demonstrated significant learning results. The design and implementation of authentic, meaningful, and manipulable visualization models is not trivial. These simulated mental models must relate to the more difficult concepts in the text, the concepts must be concrete and visualizable, the models must lend themselves to multiple appropriate distracters in the on-screen toolbox (for scoring purposes), and the correctly chosen elements from the tool box must fit into only one place on the screen for deterministic scoring. Making all these aspects come together in a seamless, gamelike environment is time consuming and certainly the most expensive component in our application.

The computer can be a powerful medium with which to train visualization. The process has been broken down into several stages: introduction/training, demonstration, practice, and final assessment. Introduction and training occur in the first *3D-Readers*' training module. Students learn why visualizing is important and some tips on how to do it. Within each interactive strategy is always an animated demonstration on what is required. Practice occurs within the interactive strategy and also via the static Brain Cloud. Finally, all interactive strategies are scored, and immediate feedback is provided. These stages are similar to the stages in McNamara et al.'s iSTART training sections: introduction, demonstration, and practice (see chap. 16, this volume).

How Do Visual and Verbal Strategies Compare to Each Other? Johnson-Glenberg (2000) described a metacognitive intervention study with 45 adequate decoders but poor comprehenders. In this study, third- through fifth-graders (average chronological age: 9;11) were randomly assigned to three groups. Two groups were instructed in different comprehension strategies. The first was the package of verbal strategies from RT (Palincsar & Brown, 1984), and the second was a visual package called *Visualizing/Verbalizing*, created by Bell (1986, 1991). The third group was composed of untreated controls. Visualizing/Verbalizing is a primarily visual strategy in which readers learn to mentally create and describe "movies in their heads" as they read. After 10 weeks of training, the two experimental groups demonstrated significant gains on several measures related to reading comprehension compared with the untreated control group. The experimental groups constructed better answers on both implicit and explicit questions and demonstrated gains in word recognition skills. Thus, strategy training aided the readers and both verbal and visual strategies were effective.

OTHER ELEMENTS UNIQUE TO 3D-READERS

The Texts

Because *3D-Readers* is supporting the instruction of lifelong strategies, texts needed to be written that spanned several genres and would lend themselves to the inclusion of the embedded strategies. The texts followed very specific guidelines: They had to be engaging in an age-appropriate manner; be synchronized with the National Science Education Standards, American Association for the Advancement of Science, California, Texas, and Wisconsin middle school content standards for science and language arts; describe content in a manner amenable to the graphical mental model strategies; and be approximately 2,000 words so the student could be finished in a half-hour time period. "Tough to teach" topics were chosen. The texts in these studies had an average Flesch–Kincaid Readability score of beginning of 5th grade. The leveling was then confirmed with several middle school teachers. The texts were written with a mixture of narrative and expository elements; a twin brother and sister humorously interact as they learn science lessons.

The content area of science was chosen for the first software release because science literacy represents a particular weakness in the American educational system (Third International Mathematics and Science Study Report, 1999). In addition, expository text is often more difficult to understand and retain. Narrative texts more closely resemble oral storytelling, and the knowledge of narrative story structure is acquired by most children before reaching school age (Stein & Glenn, 1979). The narrative style is also more accessible because it depicts event sequences that people in a culture directly enact or experience (Graesser, Golding, & Long, 1991). Narrative may be useful as a bridging vehicle for students attempting the more difficult task of expository/informational comprehension. The working hypothesis behind the choice of mixing narrative and expository styles was that encountering the familiar structure of narrative would ameliorate the increase in cognitive load typically associated with processing novel, scientific content.

The majority of the research comparing narrative and expository texts has focused on memory for the content rather than learning. In addition, most studies do not control for content across genre type. In general, memory for narrative content is better than memory for expository content (Graesser, Hauft-Smith, Cohen, & Pyles, 1980; Kintsch & Young, 1984; Wolfe & Mienko, 2006; Zabrucky & Moore, 1999). However, Roller and Schreiner (1985) found no differences between narrative and expository texts on multiple-choice items and summaries postreading. Wolfe (2005) recently found that learning and recall depended on both genre and prior knowledge, such that the higher knowledge readers benefited more from the expository text.

The majority of these studies have been done with college students and not children tackling expository for the first time. However, McNamara, Floyd, Best, and Louwerse (2004) found with third-grade readers results similar to those reported by Wolfe (2005); comprehension of science text depended on prior knowledge; moreover, comprehension of narratives depended more on reading skill.

Prompts to Reread

Mature (good) readers use the strategy of rereading (Pressley, 2000). Memory for text has been shown to improve with rereading (Amlund, Kardash, & Kulhavy, 1986; cf. chap. 7, this volume). Millis, Simon and tenBroek (1998) posited that rereading facilitates comprehension because it allows readers to complete processes that produced less-than-ideal outputs from prior reading. Poor readers are often disinclined to reread text. They may not understand that the goal of reading is to extract meaning. Poor comprehenders may also be unaware that they are not comprehending, or they may not know that the rereading strategy even exists. In the first study, rereading is examined to ascertain whether practice on other higher order, metacognitive reading strategies might affect the strategy of rereading.

In *3D-Readers*, students are always prompted to reread before working on the visual or verbal strategies. In the first study, rereading was operationalized as *ScrollBacks*. Readers would click on one of two buttons to scroll back through a section of text. Only four lines of text were visible at a time. In a window in the center of the screen, four lines were legible at a time. The text before and after the reading window was masked with case-sensitive *X*s, and the background was grey. On the right side of the screen were four buttons. The bottom two buttons scrolled the reader forward, and the top two scrolled the reader back through one section of text. There were five sections per text. The double upward-pointing arrow scrolled readers up five lines of text. The single upward-pointing arrow scrolled readers up one line of text. (This somewhat constrained reading interface was used only in the very first study and has since been altered to a more natural 12-line window.)

Writing Constructed Answers: Final Open-Ended Questions

After finishing the postreading vocabulary test, students answered eight open-ended questions in Studies 1 and 2 and six questions in Study 3. Question answering is one of the effective strategies specifically mentioned by the NRP (2000). Students generate (construct) their responses and these are sent through a spell-checker with three levels of assessment. Final content assessment scores are received in less than 2 seconds.

It is appropriate to offer some thoughts here on response scoring. Most comprehension assessment relies on answering multiple-choice questions. This is a limited knowledge assessment paradigm. In addition to adding a level of chance (usually 25%–33%) to a student's score, it relies on a rigid, one-correct-answer template. This format erroneously sends students in search of the "one true meaning" of the text (Pearson & Hamm, 2001). The final open-ended questions in *3D-Readers* require students to construct more authentic answers. The system uses HEMA to score a reader's generated questions and constructed answers. It is important to note that even though HEMA represents the vanguard of computerized automated scoring, a computer will never be as flexible and creative as a human scorer. No claims are made that the processing in HEMA is isomorphic to human cognition. HEMA is a tool designed to mimic, as closely as possible, human scoring of written language (see also chap. 16, this volume).

3D-READERS' USAGE SECTION: ELEMENTS IN ORDER OF OCCURRENCE

1. *Prior knowledge*. Students are first prompted to answer an open-ended question on the main topic; for example, "I bet you have eaten a lot of strange things in your life. Would you ever eat bacteria? Why or why

not?" This response is not scored. A student's prior knowledge semantic space is considered unconstrained, and HEMA would not be highly accurate in scoring the quality and extent of students' prior knowledge. In addition, students should not be made to feel they are in test mode with the first task. The system saves students' responses in the database for the teacher to review.

2. *Prereading vocabulary*. Students answer seven multiple-choice vocabulary items. They do not receive performance feedback at this point. On the final scores page, students are shown how their performance from pretest to posttest compared.
3. *Text reading*. Texts are split into six sections interrupted with two question-generation strategy prompts, two visual strategies, and one Brain Cloud.
4. *Embedded visualization strategy*. The visualization process has been broken into three stages that encourage an increase in generation and introduce assessment. In each module, these stages consist of one Brain Cloud and two interactive, scorable strategies.

Stage I—Static Brain Cloud. Students are asked to visualize one element from the text. This visualization happens relatively early in the text. After pausing for a mandatory 5 seconds, students are shown an example of an image.

Stage II—Visualization–demonstration. Embedded in the text are two interactive visual strategies. These are opportunities for the students to manipulate objects on screen and build examples of their ongoing mental state of comprehension. The theory is that the strategies facilitate the creation of a mental model. Because every visual strategy is unique, each begins with a demonstration of what is expected. For example, the computer will animate a drag-and-drop sequence so the student understands the requirements of the task. Students are allowed to review the demo as often as they like until they feel ready for the assessment component.

Stage III—Visualization, interactive and assessed. After the demonstration stage, the students are prompted to finish the strategy for assessment. Students drag icons or words from the toolbox and complete a final representation of their mental model, and they then submit the model for automatic scoring. This uses a module-specific rubric. After three submissions, if there is still an incorrect configuration in the student's model, the system will build the correct model for the student. Final scores and number of tries are maintained in the database. A sequence of sample screens assessing comprehension of a text that described rates of bacterial growth in an experimental kitchen is depicted in

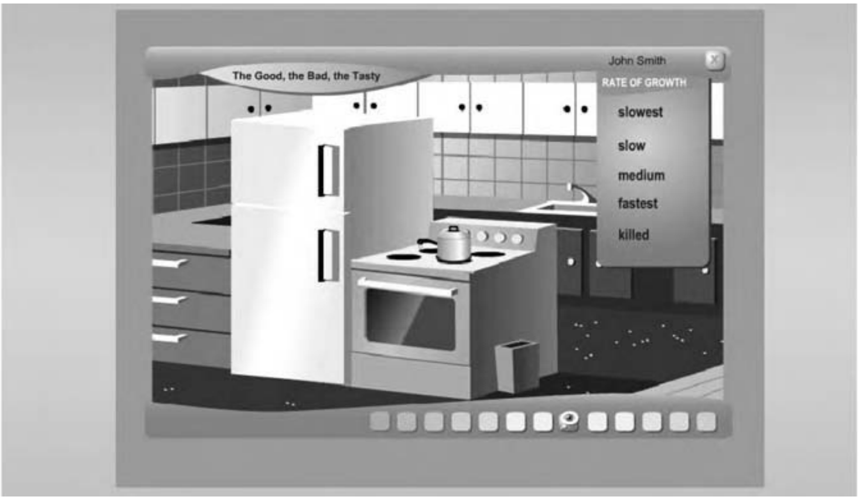


Figure 12.3. Screen shot of visual strategy in text about bacterial growth.

Figures 12.3 and 12.4. Users are instructed to drag the rate of growth from the toolbox on the right to the appropriate interactive hot spots in the kitchen. Figure 12.4 shows an example of a user who received 100% on the first try.

5. *Embedded strategy*. The embedded verbal, metacognitive strategy is question generation. Students are prompted twice to create questions. Tovani (2000) asserts that readers who ask questions improve their comprehension by interacting with the text and remaining focused on it. Asking questions fosters curiosity because readers are forced to continue reading to answer their questions; clarify information in the text; and, finally, go beyond literal meanings and to begin the first step in the processes of deduction or inference. Questions help keep readers focused and mindful of their comprehension. Asking and answering questions prompts readers to "unpack" their thinking (see also chap. 11, this volume). Quality of question generation is scored with HEMA. Students are trained to create questions that result in multiword answers. They also learn that effective questions require deeper processing of the text or even inferencing to answer. These questions often begin with "how" or "why." Examples are given of questions that are variously effective: not very effective—"Did you like the story?", more effective/good—"What part of the eye do you see when someone has 'red-eye' in a photograph?", and most effective/excellent—"How does someone get 'red-eye' in a photograph?"

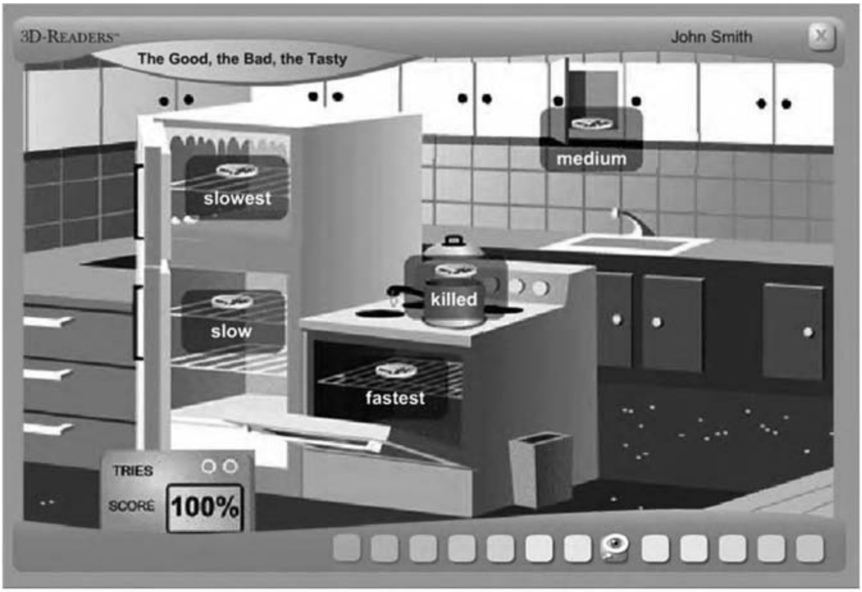


Figure 12.4. Screen shot of bacterial rates of growth correctly placed on first attempt.

6. *Postreading vocabulary*. After reading the entire text, students answer again the seven vocabulary items. This time, they receive immediate feedback on the correctness of each choice.
7. *Answering final questions*. Students are asked six open-ended questions. The answers are scored in the range of 0 to 6. If students receive a 4 or 5, they are asked if they would like a chance to try again and resubmit. If students receive a 3 or lower, they are automatically taken back to the place in the text where the answer may be found. The scores for both attempts are kept in the database. Teachers are able to check and see which students are not making improvements with rereads. A sample screen that a student might see after a second submission is depicted in Figure 12.5. The question was: "What was the experiment testing?" The constructed short answers allow students to practice writing skills as the automatic scorer assesses elements of writing quality.
8. *Final scores page*. On the final scores page, students see all the scores associated with their responses in the 0–6 format. On the final screen, scores are also displayed in the more traditional letter grade format, for example, 0 or 1 = F, 2 = D, 3 = C, 4 = B, 5 = A, and 6 = A+.

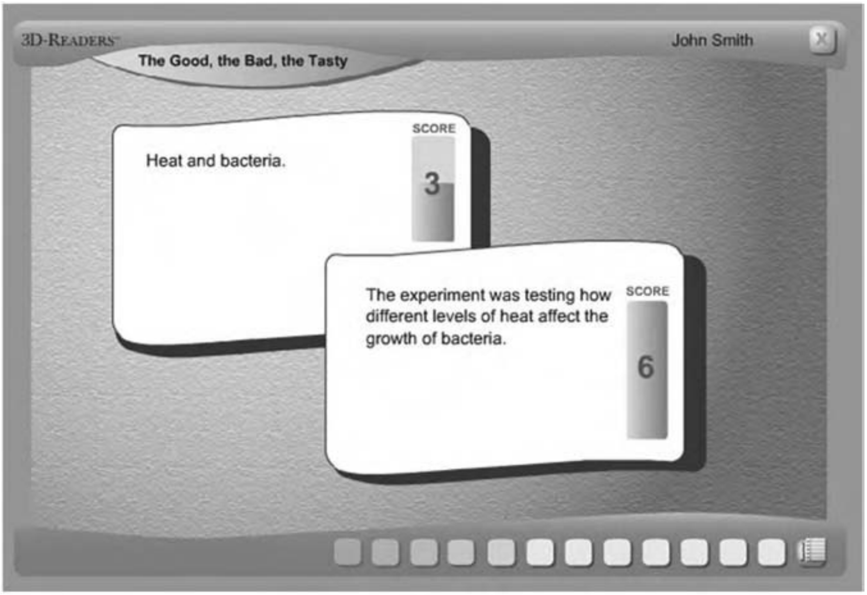


Figure 12.5. Screen shot of a second constructed response submission for the question "What was the experiment testing?"

THE STUDIES

The following studies focus on several key components in the *3D-Readers* program: vocabulary gains, metacognitive strategy use, question generation skill, rereading, and answering open-ended questions. All three studies were delivered over the Web. The first two included on-site administration and guidance by the author and several trained researchers. The third study represented natural classroom situations wherein no direct oversight was provided and teachers used the system at their discretion.

3D-Readers was created as a research-based, commercializable product with Small Business Innovation Research grant funds. The primary goal was that of increasing reading comprehension in users. I hope my colleagues and I are not guilty of the "kitchen sink" approach to intervention design as mentioned by Graesser (chap. 1, this volume). In any case, my colleagues and I consider ourselves disciplined plumbers, and the strategies and techniques included are based on solid research and amenability to computer delivery. The first study uses a manipulated control condition. It asked whether embedding the metacognitive strategies (question generation and visualization) will result in comprehension gains over and above those witnessed in the control condition without the metacognitive strategies. Table 12.1 should clarify the elements that differed between conditions.

STUDY 1: POOR COMPREHENDERS

Method

Participants. The participants were 20 poor comprehenders from an urban middle school in Madison, WI. Participants had an average percentile score of 33.26% (*SD* = 17.30) on the state's standardized reading test (TerraNova Reading and Language Arts Test; CTB/McGraw-Hill, 1997). The reading specialist recruited participants using prior contact with the students and/or referrals from teachers. The ethnicity breakdown was as follows: 45% African American, 40% White, 10% Asian , and 5% Hispanic. The sample was 55% female. Six participants were in the seventh grade, and 14 were in the sixth grade, with an average chronological age of 12;5 (range: 11;10–13;6).

Design. Because of resource constraints, and to lessen the statistical variance, a within-subject design was chosen. There were eight sessions in total; the study was designed to assess whether metacognitive strategies increased comprehension. In cognitive interventions, the control condition must always come first. Thus, the first three sessions represent the control condition. During these sessions, readers unscrambled anagrams in lieu of *3D-Readers*' higher order verbal and visual metacognitive strategies. Readers were always warned before an anagram appeared and prompted—by means of voiceover and text—if they wished to go back and reread anything. This is similar to the prompting in the experimental sessions before strategies. The purpose of these sessions was to serve as a lexical (nonmetacognitive) task and to equate all sessions for time on task.

The fourth session was a training and practice session in which the readers were introduced to the strategies (this session's data were not analyzed). Sessions 5 through 8 represented the experimental condition, and these sessions included the metacognitive verbal and visual strategies. In both conditions, no one text followed another more than one time. The control texts were written at the same time and in the same style as the experimental texts. Texts were randomly selected to become either experimental or control.

Prior topic knowledge and prereading vocabulary scores were not statistically significantly different between the two conditions. (In Study 1 only, prior knowledge was assessed using a 1–7 self-report scale.) Sessions lasted approximately 30 minutes. The entire study lasted 2 weeks.

TABLE 12.1
Differences Between the Control and Experimental Conditions

Control condition:	*Amount per module*
Control condition:	
Unscramble word anagrams (this was unique to control)	5 times
Prior-knowledge activation	1 prompt
Vocabulary	Same 7 pretest and posttest items
Prompt to reread	5 times
Construct answers (no feedback)	8 times
Experimental condition:	
Included the last four elements from above, and training in the following metacognitive strategies	
Verbal	
What makes a good question	1 training module only
Create questions (feedback)	2 times
Construct answers (feedback)	8 times
Visual	
Why visualizing is important	1 training module only
Brain Cloud (image one static element)	1 time
Construct multi-element mental model (feedback)	2 times

Results

Comprehension: Constructed Final Answers. Once readers exited the text, they could not go back to reread. Students typed in constructed answers to eight final questions for each text. These were scored using two methods: (a) human experts and (b) HEMA. The average human score significantly correlated with the HEMA scores ($r = .79$, $p < .001$).

To address the hypothesis of whether embedding strategies aided in comprehension, a paired *t* test compared final constructed answers during the control condition with the experimental condition. Using the human expert scores, the analysis revealed a significant mean difference of 7.84, $t_{(19)} = 3.14$, $p = .005$; using the HEMA scores, the analysis revealed a significant mean difference of 8.52, $t_{(19)} = 4.60$, $p < .001$. The relevant statistics, including Cohen's *d* as the effect size, are listed in Table 12.2. Thus, participants did significantly better on the texts that encouraged and allowed them to work on metacognitive strategies.

One obvious question is: Does comprehension simply improve with time by means of a practice effect, i.e., is there an ongoing linear increase such that readers who work through more texts merely continue to improve their scores with time and this favors the later experimental condition? Graphs reveal that within each condition there are slightly negative slopes associated with time. To answer this question inferentially, a within-subject growth curve analysis was conducted to compare the changes in comprehension across the control sessions with the changes in comprehension across the experimental sessions. The interaction of time and condition on comprehension was quantified for each student by an individual regression coefficient. A paired *t* test was then used to determine if the means of the interaction coefficients differed from zero. The *t* values were nonsignificant, less than 0.90. Thus, there is not much evidence for a linear practice effect that would explain the experimental condition's overall superiority in comprehension, although we do note that the power in this analysis was low.

Comprehension: Vocabulary Gains. Seven words from each text were chosen as the vocabulary items. These were agreed to be the most irregular, ambiguous, or age-infrequent words from each text. The relevant statistics are presented in Table 12.3.

A paired *t* test revealed that the average vocabulary gain from prereading to postreading in the control condition was significant (mean gain = 16.43%), $t_{(19)} = 4.57$, $p < .001$. The average vocabulary gain in the experimental condition with strategies was also significant (mean gain = 10.79), $t_{(19)} = 4.22$, $p < .001$. The interaction between gain and condition, that is, the difference in post-reading minus pre-reading scores between the two conditions, was not significantly different, $t_{(19)} = 1.48$, *ns*. We had predicted the experimental group would demonstrate greater gains on vocabulary in the postreading condition, but this was not the case. This may be because the anagram task encouraged readers to focus on the lexical and orthographic levels of the text. Anagrams (and the vocabulary words) were typically the harder words from the text. Thus, extra attention may have been given to the new, more difficult, and often irregular words during reading in order to unscramble them later. This attention could have aided postreading vocabulary definition, as the vocabulary words were also usually the more difficult and irregular words from the text.

Rereading as Assessed With ScrollBacks. ScrollBacks were computerized tallies of elective rereading. The majority of ScrollBacks were 0 and 1 (no scrolling to reread, or scrolling up only one line). Because some readers did use the double arrow (which scrolled up a block of five lines of text), we had several outlier high scores, ranging from 34 to 64. To normalize the

TABLE 12.2

Human and High-Dimensional Expert Match Algorithm (HEMA) Means, Standard Deviations (in Parentheses), Grand Means , Effect Sizes and Correlations by Session and Condition for Comprehension Assessed with Constructed Answers

Scoring	*Control*			*Grand Mean (GM)*	*Experimental*				*GM*	*Effect size*	*r (GM) Human/*
type	*1*	*2*	*3*	*(1-3)*	*5*	*6*	*7*	*8*	*(5-8)*	*(GM)*	*HEMA*
Human	46.58	36.84	36.37	40.33	51.17	47.91	44.31	46.61	48.18	.45	Control
	(25.19)	(20.99)	(22.07)	(19.58)	(21.92)	(20.73)	(21.59)	(21.17)	(15.69)		.81**
HEMA	28.51	24.71	28.77	27.25	35.95	34.83	36.34	35.92	35.76	.67	Experimental
	(19.73)	(8.51)	(10.17)	(13.59)	(11.96)	(12.57)	(10.29)	11.86)	(11.67)		.76**

Note. From "Web-Based Training of Metacognitive Strategies for Text Comprehension: Focus on Poor comprehenders," by M. C. Johnson-Glenberg, 2005, *Reading and Writing: An Interdisciplinary Journal, 18,* p.769. Copyright 2005 by Springer Science and Business Media, Springer/Kluwer Academic Publishers. Adapted with Permission.

**$p < .01$.

TABLE 12.3
Vocabulary: Percentage Correct, Means, t Tests, and Effect Sizes

	Prereading		*Postreading*			
Conditions	*M*	*SD*	*M*	*SD*	*t test (pre to post)*	*Effect Size*
Control (1–3)	56.39	16.47	72.73	14.93	$t_{(19)} = 4.57, p < .001$	1.04
Experimental (5–8)	60.22	17.39	71.01	17.08	$t_{(19)} = 4.22, p < .001$	.63

Note. From "Web-Based Training of Metacognitive Strategies for Text Comprehension: Focus on Poor comprehenders," by M. C. Johnson-Glenberg, 2005, *Reading and Writing: An Interdisciplinary Journal, 18,* p. 771. Copyright 2005 by Springer Science and Business Media, Springer/Kluwer Academic Publishers. Adapted with Permission.

distributions, 1 was added to each score, and then all scores were natural log transformed. Thus a normal distribution was approximated, and 0 remained 0, 1 became 0.69, and 64 became 4.17. The conditional means and correlations between ScrollBacks, postreading vocabulary, and constructed answers are listed in Table 12.4.

We predicted there would be significantly more ScrollBacks in the experimental condition after readers started to internalize the strategies, and the analyses did reveal an increase, paired $t_{(19)} = 2.16$, $p = .04$. It is interesting that, on average, the participants who were scrolling back more often were the relatively poorer comprehenders. The correlation between the difference in conditional ScrollBacks (i.e., control vs. experimental) and participants' earlier comprehension score was –.66 ($n = 20$, $p = .002$).

Discussion

Comprehension. This study demonstrates that scores on the most sensitive comprehension measure—constructed answers to final questions—were significantly higher in the experimental sessions, with strategies, than in the control sessions. Apparently, the use of higher level verbal strategies and the addition of the visual processing strategies increased text comprehension for these struggling middle school readers.

Because the verbal and visual strategies were not separately administered, the question of which strategy affected the greatest change cannot be addressed. However, many researchers support the use of multicomponential reading programs (NRP, 2000; Pressley, 2000). It may well be the case that the majority of proficient readers use both verbal and visual processes during situation or mental model creation. This is a view supported by Paivio's (1986; Sadoski & Paivio, 2001) dual code theory.

TABLE 12.4
Means, and Correlations Between Relevant Reading Variables

	Control		Experimental		Correlations					
	M	SD	M	SD	Vocabulary control (post)	Vocabulary Experimental (post)	Human Answer control	Human Answer Experimental	HEMA Answer control	HEMA Answer Experimental
ScrollBack[a]										
LogN	.96		1.35*		.01[c]	−.69**				
(0 – 4.17)	(.69)		(.95)[b]				.21	−.45*	.28	−.12

Note. HEMA = High-dimensional Expert Match Algorithm. From "Web-Based Training of Metacognitive Strategies for Text Comprehension: Focus on Poor comprehenders," by M. C. Johnson-Glenberg, 2005, *Reading and Writing: An Interdisciplinary Journal, 18,* p. 773. Copyright 2005 by Springer Science and Business Media, Springer/Kluwer Academic Publishers. Adapted with Permission.

[b] ScrollBack paired *t* test, $t_{(19)} = 2.16$, p = .04.

[c] The most relevant vocabulary score, which is postreading, is reported. In addition, the vocabulary control correlation with vocab postreading minus vocabulary prereading difference was $.32_{(Ns)}$. The vocabulary experimental correlation with vocabulary difference was .48 ($p = .03$).

[a] ScrollBack is natural log transformed. The numeral 1 was added to each score; without transform, the raw control mean was 4.65 ($SD = 5.75$), raw experimental mean was 8.07 ($SD = 10.42$).

* $p < .05$; ** $p < .01$.

Dual Code Theory. Dual code theory posits three levels of processing or meaning for both the verbal and visual codes. The first level is the *representational level*, which involves the initial activation of one or both code systems. The structure at this level can be "described as the availability in memory of modality-specific logogens and imagens (neuronal structures)" (1986; Sadoski & Paivio, 2001) dual code theory. At the second level are *referential connections*, which operate between visual and verbal systems. At the third level are *associative connections*, which operate within systems connecting imagens and logogens to one another. If comprehension is conceived of as a pattern of neural activation composed of both verbal and visual elements, then a pattern that is both highly activated and veracious represents "good comprehension." *3D-Readers* may have increased text comprehension for two reasons. First, question generation, by activating the verbal code, may force readers to review current knowledge and ascertain where their knowledge structures are incomplete or fuzzy. Question generation and answering open-ended questions at the end of the text would certainly activate both representational and associative links in the verbal system. Second, the visual strategy that entails building a mental model onscreen may activate all three levels of representational, associative, and referential links in both verbal and visual systems. Aspects of the text are turned into imagery, concepts from the text are repeated, and then readers manipulate and verify where icons should be placed on screen. Activating all three levels and stimulating communication between and within the two verbal and visual systems represents powerful across-the-board cognitive processing. This is the sort of deep processing to which researchers refer when they say that readers need to be actively engaged with the text (chap. 3, this volume). The more practice in effortful cognitive processing that poor comprehenders receive, the more proficient they should become at activating these processes on other texts and in other literacy situations.

Vocabulary Gains. Although there were no significant conditional differences on vocabulary gains, readers did show consistent overall gains in vocabulary from prereading to postreading. This supports the tenet that embedding definitions of new words close to their first exposure significantly increases understanding of those words by posttest.

ScrollBacks. The analysis of ScrollBacks was edifying. Rereading has been shown to improve comprehension and metacomprehension accuracy in college-age students (Rawson, Dunlosky, & Theide, 2000). On average, the readers in this study used significantly more ScrollBacks in the experimental condition. However, it was the relatively poorer comprehenders who were using the technique more often than the relatively better comprehenders.

Perhaps because the poorer comprehenders came to realize, through the system's feedback, that they were struggling and should avail themselves of the strategy. Perhaps by integrating the system's immediate feedback and their own growing metacognitive awareness, the poorer comprehenders began to more consciously, or at least electively, use one of the new tools available to them. Rereading is important because Haenggi and Perfetti (1992) demonstrate that, on average, college readers show equal benefits in a comparison among three different types of text reprocessing strategies: (a) rereading, (b) rewriting notes, or (c) rereading notes. Taking notes would normally be considered a more effortful strategy and, thus, more beneficial; however, simply rereading the text increased college students' comprehension scores in equal measure.

The next two studies represent pilot studies on a new version of *3D-Readers* (accessible at http://www.neuronfarm.com). The interface, the prompt for prior knowledge, and the HEMA algorithm have been modified and improved. The goals of these studies were to replicate vocabulary gains, assess changes in metacognitive strategy use, and observe use in the real world.

STUDY 2: POPULATION WITH LEARNING DISABILITIES

Method

Participants. This pilot study included 11 third- through eighth-graders in a private school for students with learning disabilities. Ten were male, and the sample was 97% Caucasian. The 2 third-graders were considered gifted and read above their grade. The other participants were in the sixth, seventh, and eighth grades. Half of them read "on level," according to the reading specialist, and the other half read one to two grade levels below. All participants were diagnosed with attention deficit disorder/attention deficit hyperactivity disorder (ADHD). If students normally took medication, then they were on medication at the time of the study.

Design. Due to the small sample size, this was a within-subject design. There were six sessions; this study was specifically designed to hone the HEMA algorithm, gather further statistics on vocabulary gains, and assess whether middle school readers would self-report more metacognitive strategy use after going through the intervention.

Materials. The six original texts were written at a beginning of fifth-grade decoding level on the Flesch–Kincaid scale. Before beginning the study, participants filled out the Metacognitve Awareness of Reading Strategies Inventory

(MARSI; Mokhtari & Reichard, 2002), which was designed for adult readers. The measure was abbreviated for this study and included 17 items (of the original 30) that were most appropriate for the study's younger population, for example, "I ask myself questions I like to have answered in the text."

Results

HEMA. Using a revised version of HEMA and including a state-of-the-art spell-checker, the correlation between blind human scorers and HEMA increased to .86 ($p < .001$).

Vocabulary. As in the first study, there were seven vocabulary items tested before and after reading. The prereading vocabulary score ($M = 65.80$, $SD = 19.11$), was significantly less than the postreading vocabulary score ($M = 81.45$, $SD = 15.49$), paired $t_{(10)} = 4.74$, $p = .001$. On average, students gained 15.65% ($SD = 10.95$) on postreading vocabulary; the gains at Time 6 compared to Time 1 increased marginally by 13% (paired t, $p = .09$).

MARSI. In the preintervention administration of the MARSI, participants scored 49.45 ($SD = 11.22$), and in the postintervention administration they scored 54.73 ($SD = 12.64$), paired $t_{(10)} = 2.28$, $p = .045$. This suggests that the *3D-Readers* intervention produced significant gains in reported metacognitive strategy use.

Discussion

This population is notoriously difficult to keep on task with hard copy materials. However, one study (Ford, Poe, & Cox, 1993) showed that the attention of teenage boys with ADHD increased when using software with a game format when animation was not excessive. Straight tutorial animation did not appear to hold attention as well as the games did. (The difficulty, format, and content of the various software were additional factors affecting the nonattending behaviors of their participants with ADHD.)

During Study 2, the computer laboratory instructor and principal commented several times how impressed they were that all the students maintained interest in reading with *3D-Readers*. Because all of the students were slower readers, it took them on average 1 hour to complete a module, rather than the usual 30 minutes observed in previous smaller pilot studies with on-level readers. Nonetheless, the students were very persistent, and the majority reported that they enjoyed it.

On average, impressive vocabulary gains were seen again. These gains replicate the findings in Study 1 and demonstrate that highlighting the vocabulary words and embedding the definitions close to the words is a valid method with which to teach the meaning of new words.

The MARSI gains were interesting, because it is unusual to see changes in metacognition in such a short time (Meloth, 1990). There was no prompting when the students retook the MARSI to remember what they had just learned in the intervention. However, social desirability effects must always be taken into account when administering self-reports of this nature. Although, the students were probably being thoughtful about the measure because at posttest, they scored a mean of 49 on a test with a ceiling of 85.

STUDY 3: SUMMER SCHOOL—A NO-OVERSIGHT EXPLORATORY STUDY

The summer school study represents *3D-Readers* in the real world. Summer school teachers logged on and used the software with no oversight. Teachers were allowed to assign the four modules in any order, and experimenters were not on site to verify correct usage of the program. For these reasons, this study should be considered exploratory. Seven teachers signed up for the free beta test summer school program in Wisconsin, California, and Texas. Of these original seven, five had students who went through the entire set of four modules, and only the data from these five teachers were analyzed. In addition, teachers were asked to fill out a short demographics survey.

Method

Participants. Of the original 95 students signed up, only 37 finished all four modules. There were 22 males (59%) in the sample. These students were in summer school, so it might be assumed that they were having some difficulty keeping up in school. Teachers did not supply us with standardized reading scores or racial information, but they often described the students as having "low motivation." Participants ranged from fourth- to seventh-graders (grade just completed). Approximately 17% were English language learner students.

Design. This was not a controlled study; instead, our goal was to observe how teachers would use the software without constraints and to analyze question generation. Teachers had 2 months in which to log on and go through the four modules in any order they chose. It was emphasized that *The Wave*, the training module, should come first. Apparently, all teachers chose the default order and students went through *The Wave* (how to measure a wave); *The Good, the Bad, and the Tasty* (bacteria and experiments); *Shameka Sees the Light* (color as wavelengths); and *Bear Necessities* (genetics and survival), in an invariant order. Because the order of the texts was not randomized, effects across sessions need to be interpreted with caution.

Results

Over the four sessions, many elements within *3D-Readers* varied, for example, the texts, the final questions, the visual strategies. The only scored variable that remained consistent throughout the study was the question generation prompt. Thus, this study focused on changes in quality of self-generated questions. In addition, to remain parallel with Studies 1 and 2, vocabulary gains and final comprehension question analyses are also reported.

Vocabulary. Pre- and postreading vocabulary scores were gathered on 145 modules. These were aggregated by participant: prereading mean = 67.83 (SD = 19.44) and postreading mean = 81.01 (SD = 16.71). A significant average gain of 13.19 was observed, paired $t_{(36)} = 5.67$, $p < .001$.

Comprehension. Participants' constructed responses were all scored by HEMA. *The Wave* is the shortest and "easiest" module to comprehend, and thus we would expect the highest scores on it. In addition, the users' general consensus has been that *Bear Necessities* is one of the more difficult modules, with tougher questions. Because order of modules was not randomized, these confounds have not been controlled for, so it should be assumed that the first module was the easiest and the last module was the hardest. A review of the raw, unadjusted scores reveals a slight negative slope in comprehension scores over time (Session 1 = 62.02, Session 2 = 58.61, Session 3 = 54.74, Session 4 = 52.35). However, if pretest knowledge of vocabulary words is controlled for, this slope reverses sign (but not significantly). A second analysis was conducted, controlling for pretest vocabulary (i.e., prior knowledge). The adjusted comprehension scores are reported in Table 12.5, which lists the means, standard deviations, and Cohen's d effect sizes between Sessions 1 and 4.

Question Generation. The one variable that remained constant across modules was the prompt to create the question "Pretend you are a teacher. What question would you ask to check your students' understanding of the story so far?" The hypothesis is that linear gains should be seen over time on question generation quality because students are receiving numerical feedback

TABLE 12.5
Adjusted and Unadjusted Means, and Effect Sizes (ES) of Key Variables

	Session 1		*Session 2*		*Session 3*		*Session 4*		
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*ES between sessions 1 & 4*
Answer final comp. questions (adj.)	57.58	2.00	57.07	2.56	58.00	2.33	58.10	2.74	.22
Answer final comp. questions (unadj.)	62.02	15.12	58.61	17.67	54.74	14.37	52.35	15.93	.62
Letter grade Equivalent	C		C		C		C		
Question generation	47.65	15.99	60.14	14.99	71.46	17.84	67.81	22.93	1.04
Letter grade Equivalent	D		C		B		B		

Note. All variables were scored with High-dimensional Expert match Algorithm. Comp. = comprehension.

after each submission (two submission opportunities per module). The last two rows in Table 12.5 list the study's means, standard deviations, and the effect sizes. Readers' question generation scores increased significantly over the course of the study. A repeated measures analysis revealed that means across the four sessions differed significantly, $F_{(3, 111)} = 19.62$, $p < .0005$, with the best performance being almost two grade levels better than performance on the first session. Planned contrasts revealed a significant linear trend, $F_{(1, 37)} = 31.3$, $p < .0005$, across sessions. In addition, the quadratic effect was significant, $F_{(1, 37)} = 11.33$, $p = .002$, suggesting that performance leveled off after the third session. A longer term study is needed to further understand this function over time. Obviously, 71% is not the ceiling, and further gains might be expected with increased training.

GENERAL DISCUSSION

These three studies present compelling evidence for the benefits of using computers to train young readers in metacognitive strategies and allow practice with them. Short-term use of the program resulted in significant gains in several key reading areas. The ultimate goal of the application is to teach readers how to monitor comprehension and then use appropriate repair strategies. Monitoring can be achieved by means of question generation (with additional internal effort expended to then answer the generated question), assessing vocabulary gains, and constructing answers at postreading. The system encourages the use of the several repair strategies, such as rereading and visualizing. The five most important findings in the three studies can be summarized as follows.

Study 1, with poor comprehenders, (a) revealed significant gains on constructing answers to open-ended questions. In addition, the study revealed that poor comprehenders can (b) alter their reading processes when encouraged to go back into the text and reread.

Study 2, with students with attention deficit disorder/ADHD, revealed (c) significant vocabulary gains and (d) significant gains in self-reported metacognitive strategy use after six sessions.

Study 3, with summer school students, also revealed significant gains in vocabulary, as well as (e) significant gains in the quality of questions generated over four sessions. Question generation is one of the premier metacognitive strategies for increasing monitoring and comprehension. However, students in Study 3 showed little comprehension gains. It is impossible to know in this observational study whether this null effect was due to the confound of story order (with the most difficult text coming last) or the length of the study.

Importance of Findings

The findings are important for several reasons. First, they demonstrate that Web-based strategy training tools can significantly increase striving readers' comprehension scores. By allowing readers to create questions as they read, and to manipulate graphics on screen to create stylized simulations of their mental models, the readers' final comprehension scores increased significantly. In the most formal study, Study 1, the experimental condition participants' scores increased on average 8% beyond the control condition participants' scores. When assessing the difference between the control and experimental conditions with human-scored constructed answers, a Cohen's *d* effect size of 0.45 was found, and in the HEMA-scored condition (with smaller standard deviation) the effect size was 0.67. As a comparison, in a meta-analysis by Rosenshine and Meister (1994) on the metacognitive intervention RT, the median effect size with standardized measures was 0.32.

Second, these results are heartening given that the study lasted only 2 weeks. Duffy et al. (1987) and Meloth (1990) have demonstrated that it can take up to 16 weeks for significant higher level metacognitive differences to emerge in poor readers using hard copy materials. Third, these studies demonstrate that it is possible to move away from the multiple-choice format when testing with computers. Fourth, the new generation of formal and informal school assessments expects students to construct coherent and orthographically correct responses (short answers). Web-based tools that require constructed responses and attention to spelling and that encourage students to practice the extremely valuable skills of writing and rewriting are important additions to the educational toolbox. Fifth, the consistent gains in word learning seen in each study support the usefulness of computers in vocabulary instruction and the concept of authentically embedding definitions near the first word exposure. Sixth, reading instruction needs more empirical Web-based research. The following quote is from the NRP (2000): "Particularly striking in its absence is research on Internet applications as they might be incorporated in reading instruction" (p. 6-2).

I hope that studies like these will encourage other educational software designers to include more generative and constructive elements in their designs. Unfortunately, some of the new Web-based products still resemble scanned textbooks with multiple-choice assessments. Important research by Pearson, Peterson, Rodriguez, and Taylor (2002) lends support to the concept of allowing students to be more generative in the classroom. Those authors encouraged teachers to provide more small-group instruction, encourage pupil engagement, and use "coaching—rather than telling." They include powerful examples of coaching that are based on asking students questions. In their study, fourth-, fifth-, and sixth-grade teachers were followed, and researchers coded every 5-minute block of instruction. In hierarchical linear modeling analyses, students in classes with high levels of questioning (coaching) demonstrated a significant positive relationship on the Gates–MacGinitie test of reading. Students in classes with high levels of "telling" demonstrated a significant negative relationship on the test. We interpret coaching as a form of scaffolded generativity; it is an active process. Active processing with text is also advocated by Oakhill and Cain (chap. 3, this volume). They encourage active involvement with the meaning of a text and posit that this should help readers to both foster learning and "an attraction to reading." As seen in Study 3, questions of significantly higher quality were generated by the students themselves in only four sessions.

Shortcomings and Future Directions

There are several problems inherent to studies of practical interventions. As Graesser (chap. 1, this volume) notes, many pragmatic considerations must be taken into account when creating effective interventions for real world classrooms. Teachers will not spend precious funds on interventions designed for one-degree-of-freedom academic studies. They are, however, interested in packages containing multiple research-based strategies. The manipulation in Study 1 demonstrated that greater comprehension gains could be seen after exposure to both verbal and visual metacognitive strategies—versus a control condition with a more lexical-level strategy. I certainly do advocate future studies using large between-subjects designs to pull apart the efficacy of the strategies and to ascertain which instructional components truly benefit which types of readers. Another important question is how the hybrid text is affecting comprehension. The texts addressed scientific concepts that were introduced in an engaging, narrative structure. One hypothesis to be tested is whether the hybrid story format is more felicitous for younger readers, who might be struggling with the novel expository format, than it is for older readers, who might find it distracting. In addition, there are interactions between text type (expository, narrative, or hybrid) and metacognitive strategy (verbal, visual, or mixture) that need to be addressed.

The issue of transfer is an important one, and we did not have the resources to address it. The strategies could be "faded" toward the end of training, and performance on computer-delivered texts without embedded strategies can be assessed. In addition, a far-transfer assessment with hard copy texts may be even more ecologically valid. My colleagues and I are also actively researching the addition of more elaborative, verbal feedback on the scoring of constructed responses. It would be valuable to move beyond a numerical score and inform students why they received a certain score, as well as giving specific feedback as to how they might increase their score in the next submission. Another shortcoming was the lack of audio support for English language learner students.

Research to date supports the supposition that embedding verbal and visual strategies in Web-based instructional applications can result in significant improvements in text comprehension. Although more research is needed, the current system holds great promise for increasing comprehension strategy use in middle school readers, especially in several atypical populations.

ACKNOWLEDGMENTS

This research was supported by two grants from the National Institutes of Health Small Business Innovation Research, R43 HD040696-02A1 and R44 HD040696-02A1. I thank Arthur M. Glenberg, Gary McClelland, David Elderbrock, Ankur Malhotra, William Kunkler, and all the young readers who participated in these studies.

REFERENCES

Amlund, J. T., Kardash, C. A. M., & Kulhavy, R. W. (1986). Repetitive reading and recall of expository text. *Reading Research Quarterly*, *21*, 49–58.

Bell, N. (1986). *Visualizing and verbalizing for language comprehension and thinking*. Paso Robles, CA: Academy of Reading.

Bell, N. (1991). Gestalt memory: A critical factor in language comprehension. *Annals of Dyslexia*, *41*, 246–260.

Doctorow, M. J., Wittrock, M. C., & Marks, C. B. (1978). Generative processes in reading comprehension. *Journal of Educational Psychology*, *79*, 109–118.

Duffy, G. G., Roehler, L. R., Sivan, E., Rackliffe, G., Book, C., Meloth, M. S., et al. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, *22*, 347–368.

Duke, N. K., & Pearson, P. D. (2002). Effective practices for developing reading comprehension. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed., pp. 205–242) Newark, De: International Reading Association.

Durkin, D. (1978–1979). What classroom observations reveal about comprehension instruction. *Reading Research Quarterly*, *14*, 481–533.

Elshout-Mohr, M., & van Daalen-Kapteijns, M. (2002). Situated regulation of scientific text processing. In J. Otero, J. A. Leon, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 223–252). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Ford, M. J., Poe, V., & Cox, J. (1993). Attending behaviors of ADHD children in math and reading using various types of software. *Journal of Computing in Childhood Education*, *4*, 183–196.

Gambrell, L. B., & Jaywitz, P. B. (1993). Mental imagery, text illustrations and children's story comprehension and recall. *Reading Research Quarterly*, *28*, 265–273.

Garcia, G., Pearson, P. D., & Taylor, B. (2004, December). Presentation to the National Reading Conference. San Antonio, TX, Current State of Strategy Instruction.

Graesser, A. C., Golding, J. M., & Long, D. L. (1991). Narrative representation and comprehension. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 171–205). White Plains, NY: Longman.

Graesser, A. C., Hauft-Smith, K., Cohen, A. D., & Pyles, L. D. (1980). Advanced outlines, familiarity, and text genre on retention of prose. *Journal of Experimental Education*, *48*, 281–290.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.

Haenggi, D., & Perfetti, C. A. (1992). Individual differences in reprocessing of text. *Journal of Educational Psychology*, *84*, 182–192.

Hegarty, M., Narayanan, N. H., & Freitas, P. (2002). Understanding machines from multimedia and hypermedia presentations. In J. Otero, J. A. Leon, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 357–384). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Johnson-Glenberg, M. C. (2000). Training reading comprehension in adequate decoders/poor comprehenders: Verbal vs. visual strategies. *Journal of Educational Psychology*, *92*, 772–782.

Johnson-Glenberg, M. C. (2005). Web-based training of metacognitive strategies for text comprehension: Focus on poor comprehenders. *Reading and Writing: An Interdisciplinary Journal*, *18*, 755–786.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W., & Young, S. R. (1984). Selective recall of decision-relevant information from texts. *Memory & Cognition*, *12*, 112–117.

Levin, J. R. (1973). Inducing comprehension in poor readers: A test of a recent model. *Journal of Educational Psychology*, *65*, 19–24.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge, England: Cambridge University Press.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, *86*, 389–401.

McNamara, D. S., Floyd, R. G., Best, R., & Louwerse, M. (2004). World knowledge driving young readers' comprehension difficulties. In Y. B. Yasmin, W. A. Sandoval, N. Enyedy, A. S. Nixon, & F. Herrera (Eds.), *Proceedings of the sixth international conference of the learning sciences: Embracing diversity in the learning sciences* (pp. 326–333). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McNamara, D. S., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43.

Meloth, M. S. (1990). Changes in poor readers' knowledge of cognition and the association of knowledge of cognition with regulation of cognition and reading comprehension. Journal of Educational Psychology, 82 (4), 792–798.

Millis, K., Simon S., & tenBroek J. (1998). Resource allocation during the rereading of scientific texts. *Memory & Cognition*, *26*, 232–246.

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, *94*, 2, 249–259.

Naps, T. L., Roling, G., Almstrum, V., Dann, W., Fleischer, R., Hundhausen, C., et al. (2003). *Exploring the role of visualization and engagement in computer science education*. Retrieved October 10, 2004, http://elara.tk.informatik.tu-darmstadt.de/publications/2003/wgReport.pdf

National Reading Panel. (2000). *Teaching children to read: Reports of the subgroups*. Retrieved June 5, 2001, from www.nichd.nih.gov/publications/nrp/report.htm

Oakhill, J., & Patel, S. (1991). Can imagery training help children who have comprehension problems? *Journal of Research in Reading*, *12*, 106–115.

Paivio, A. (1986). *Mental representation: A dual coding approach*. New York: Oxford University Press.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*, 117–175.

Pearson, P. D., & Hamm, D. N. (2001). *The assessment of reading comprehension: A review of practices—Past, present and future*. Technical report for RAND corporation.

Pearson, D. P., Peterson, D. S., Rodriguez, M. C., & Taylor, B. M. (2002). Looking inside classrooms: Reflecting on the "how" as well as the "what" in effective reading instruction classroom practices. *The Reading Teacher*, *56*, 270–279.

Pressley, M. (1977). Imagery and children's learning: Putting the picture in developmental perspective. *Review of Educational Research*, *47*, 585–622.

Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 545–561). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Rawson, K. A., Dunlosky, J., & Theide, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, *28*, 1004–1010.

Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology*, *80*, 16–20.

Roller, C. M., & Schreiner, R. (1985). The effects of narrative and expository organizational instruction on sixth-grade children's comprehension of expository and narrative prose. *Reading Psychology: An International Quarterly*, *6*, 27–42.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, *64*, 479–530.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, *66*, 181–221.

Sadoski, M., & Paivio, A. (2001). *Imagery and text: A dual coding theory of reading and writing*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 592–604.

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Technical report for the RAND Corporation.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing* (pp. 53–120). Norwood, NJ: Ablex.

TerraNova Reading and Language Arts Test. (1997). Monterey, CA: CTB/McGraw-Hill.

Tovani, C. (2000). *I read it but I don't get it: Comprehension strategies for adolescent readers*. Portland, ME, Stenhouse.

Wittrock, M. C. (1991). Generative teaching of comprehension. *Elementary School Journal*, *92*, 169–184.

Wittrock, M. C., & Kelly, R. (1984). Teaching reading comprehension to adults in basic skills courses (Final report, Project No. MDA 903-82-C-0169). University of California, Los Angeles, Graduate School of Education.

Wolfe, M. B. W. (2005). Memory for narrative and expository text: Independent influences of semantic associations and text organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 359–364.

Wolfe, M. B. W., & Mienko, J. A. (2006). *Learning and memory of factual content from narrative and expository text*. Manuscript submitted for publication.

Zabrucky, K. M., & Moore, D. (1999). Influence of text genre on adults' monitoring of understanding and recall. *Educational Gerontology*, *25*, 691–710.

13

Visiting Joke City: How Can Talking About Jokes Foster Metalinguistic Awareness in Poor Comprehenders?

Nicola Yuill
University of Sussex

Some children have poor text comprehension skills, despite being good decoders. Part of becoming literate involves understanding and reflecting on how text conveys meaning. This chapter addresses the role of metalinguistic awareness of meaning in explaining and in improving poor comprehension. In 2 main studies, groups of 7- to 9-year-old children were supported in discussing verbal ambiguities as a way of focusing attention on meaning and thereby improving comprehension. A piece of software, Joke City, engaged pairs of children in discussing joking riddles with plays on meaning and was shown to produce comprehension improvement on a standardized test. High comprehension improvement was associated in particular with increases across training sessions in the use of metalinguistic statements contrasting different meanings. The results demonstrate children's metalinguistic abilities, show the potential for metalinguistic training, and illustrate how children's spoken language can be used to index their metalinguistic skills.

It is increasingly acknowledged that some children have difficulty developing text comprehension skills, even if their decoding skills are good. The chapters in this volume present many useful interventions to foster comprehension skills. In particular, there is increasing emphasis on the use of metacognitive strategies: sets of mental tools that children can use strategically to improve their understanding of text. These approaches tend to focus on helping children to develop and apply articulated knowledge, such as that of typical structures of text (e.g., chap. 8, this volume) and the child's own domain knowledge (chap. 4, this volume). However, surprisingly little work has addressed the role in comprehension of *metalinguistic awareness*, the ability to adopt a reflective attitude toward the comprehension and production of oral and written language (Gombert, 1992). Being able to reflect on, analyze, and talk about language would seem to be a natural precursor to the intelligent use of any metacognitive strategy in reading: For example, how could one summarize the main ideas of a text without being able to reflect on or analyze what those ideas were?

This chapter starts from the idea that children of around 7 to 9 years of age with poor comprehension are deficient in metalinguistic awareness, in particular of the semantic aspects of text, below the level of text macrostructure, that is, being aware of how words and phrases in text convey meaning. It focuses in particular on children's awareness that words and phrases can have multiple meanings and can be ambiguous. If there is a link between this aspect of metalinguistic awareness and text comprehension, then training children's awareness of multiple and ambiguous meanings could play an effective role in fostering their text comprehension. Verbal jokes are a naturally occurring example of children's language use in which the focus is on interpreting alternative meanings. Jokes might therefore be an ideal means of focusing children's attention onto meaning and increasing their sensitivity to meaning. In each of the studies summarized herein, children's conversations about text are analyzed to throw light on children's capacity to articulate their reflections on language and to see how this capacity might index improvements in comprehension.

The chapter is organized into four parts. It begins with a brief account of the role of metalinguistic awareness in becoming literate and a review of evidence suggesting poor metalinguistic awareness in children with poor text comprehension. In the next section, I summarize a study in which teacher-led training with ambiguous sentences was used to increase sensitivity to meaning in children with poor comprehension. This study introduces the idea of looking at children's conversations about text to see how improvements in metalinguistic awareness might foster changes in text comprehension skill. The third section describes a pilot study using a software game to engage pairs of children in peer discussions of ambiguity, using joking riddles, and a more formal study of the effect of such training on comprehension and on metalinguistic awareness. In each of these studies, children's conversations provide ideas about how metalinguistic talk might foster comprehension improvement and insights into the nature of metalinguistic awareness in children who are just becoming fluent readers. In the final section, some potential theoretical and practical implications of the studies are presented.

BACKGROUND: METALINGUISTIC AWARENESS AND POOR COMPREHENSION

Many cognitive skills contribute to reading comprehension, and the ability to make appropriate inferences is fundamental. The study of children with good decoding skills but poor comprehension, in contrast to age-matched children good at both aspects of reading, has been instructive in elucidating these inferential difficulties. Cain and Oakhill (1999) investigated different reasons why poor comprehenders might fail to draw appropriate inferences. They found that these children's poor performance on inferential questions about text could not be explained purely by poor memory for the text in general, or by a lack of general knowledge needed to make inferences. In particular, some children failed to make an inference needed to answer a question even if they were shown the appropriate sentence in the text and had the necessary general knowledge. For example, children read a sentence in a story, "Then they set off for home, pedaling as fast as they could," and were later asked "How did the children get home?" As expected, poor comprehenders, in contrast to good comprehenders, often failed to answer these inferential questions correctly on first being asked. They were then directed to the sentence in the text and asked "What sort of things can we pedal?" This led to improvement in performance, but some poor comprehenders still failed to answer correctly.

Cain and Oakhill (1999) suggested that the reason for this is that poor comprehenders, unlike good ones, do not strive for coherence when they read. One source of poor comprehension might therefore be in the approach taken to reading: Poor comprehenders do not seem to pick up implicit or even explicit cues to inferences, even though they can be helped to make those inferences when given support. This is consistent with other evidence that less skilled comprehenders may focus on reading accuracy rather than on comprehension. Yuill and Oakhill (1991) reported that young, less skilled comprehenders were more likely than skilled ones to define and assess reading skill in terms of accurate decoding rather than understanding, and they tended to equate good reading with good decoding. These results suggest that it is not an inability to make inferences, but a relative insensitivity or inattention to interpretive features of text, that characterizes children with poor comprehension.

Further support for this suggestion comes from studies that use deliberately ambiguous texts as a means of focusing attention on the process of drawing inferences to extract meaning from text. For example, Yuill and Joscelyne (1988) presented skilled and less skilled comprehenders with highly ambiguous narrative texts and trained children to pick out a series of clue words, to enable them to make inferences that could disambiguate the rest of the text. For example, a story with an unspecified location mentioned a girl watching the "ground below," feeling "unsafe," and her father running toward "the tree," prompting the inference that the girl was up in a high tree. Training led to improved performance on inferences in less skilled comprehenders, suggesting that they had the ability to make inferences but, perhaps because of a lack of sensitivity to meaning, failed to use this ability appropriately.

It seems from the studies just described that less skilled comprehenders often adopt a superficial style of reading through which they are not attuned to what inferences can be made—even if they can make those inferences when prompted. These difficulties are apparent when both reading and listening to text and do not seem to be entirely a result of deficits in basic cognitive processes such as memory, vocabulary, or specific domain knowledge.

One potentially fruitful way to see this difficulty is as a failure of metalinguistic awareness. Why metalinguistic and not metacognitive? There is a substantial body of work on metacognitive approaches to reading, but the focus of that work is rather different from what is being examined here. The reading and metacognition literature focuses very much on articulating knowledge about specific strategies for different reading purposes or types of text, such as skimming, summarizing, question generation, and being aware of the extent of one's comprehension. The studies presented here are concerned less with specific reading strategies and evaluation than with the more fundamental processes involved in inferences from single words and phrases in any kind of text, written or spoken. The awareness of how text can carry meaning at this level would seem to be a prerequisite for the use of any higher level metacognitive strategy.

Metalinguistic awareness is of course intertwined with metacognitive awareness. As Francis (1999) pointed out, metacognition is a general concept that implies reflective thinking, whereas metalinguistic awareness could in theory be defined purely in terms of knowledge of specific linguistic forms such as phonemes or words and their referents. The metalinguistic awareness discussed in this chapter involves metacognitive skill inasmuch as that is defined as reflective awareness of language, in this case, paying attention during reading or listening to the explicit and implicit cues to the meaning of a text. Such awareness may or may not be metacognitive, in that a reader might direct attention to meaning but not be explicitly aware of doing so.

In the studies presented in this chapter, children were engaged in metalinguistic discussion of language, in that they were talking about constructs such as "word" and "meaning," but they did not necessarily show metacognitive awareness that this might be a useful strategy to use while reading, because they were engaged more playfully in the discussion of jokes. In this respect, children are not being trained to use deliberate strategies when reading; instead, their attention is being drawn indirectly to the interpretive aspects of text.

We have seen that poor comprehenders seem relatively insensitive to cues to make inferences, but what evidence is there that these children show poor metalinguistic awareness of meaning? One method of looking at this is to present children with ambiguous texts that have multiple meanings. A series of studies on ambiguity in joking riddles provides evidence for links between comprehension skill and metalinguistic awareness. For example, how does one appreciate the following riddle: "Why do leopards never escape from the zoo?—Because they are always spotted"? It rests on understanding that *spotted* can be interpreted as a pattern (cued by the word *leopard*), but the answer makes sense only if the word is interpreted as *seen*.

I (Yuill, 1998) assessed good and poor comprehenders on the ability to understand language ambiguity in joking riddles using two different methods: (a) retelling riddles and (b) choosing the "funny" answer to a joke question from two possible answers. The correct answer used meaning ambiguity, whereas the other did not: It was either a sensible answer to the joke question or an absurdity. For example, the foil answer to the leopard riddle might be "Because the cages are strong." Accurately retelling and correct choice of joke answers were each significantly related to comprehension skill, independently of age and decoding skill. Furthermore, poorer comprehenders were not generally poor at understanding riddles: The study compared comprehension of riddles that focused on ambiguity in meaning to riddles that focused on sound (e.g. "Where do you take a hurt wasp?—To a waspital").[1] Understanding of meaning-based riddles was significantly more closely related to reading comprehension skill than to reading accuracy, whereas performance on sound-based riddles was as closely related to accuracy as it was to comprehension. Complementary evidence from other work shows evidence of a relation between reading accuracy and sound-based riddles (Mahony & Mann, 1992). I (Yuill, 1998) also reported a pilot study using riddles as a training method, with some resultant comprehension improvement. This supports the idea that focusing children's attention on language ambiguity

[1]In the studies discussed in this chapter, I used only riddles that played on meanings: There are various ways of subclassifying these (see Yuill, 1998).

might be a means of improving their comprehension, and it led to the studies described here. Such training may operate by addressing the style of reading that Cain (1999) proposed as one cause of poor comprehension.

TUTOR-LED TRAINING IN DISCUSSION OF MEANING: A PILOT STUDY (STUDY 1)

Our first rather modest attempt to engage children in metalinguistic discussion was to develop a series of three half-hour lessons for children in small groups to discuss meaning ambiguity over the course of 2 weeks. We wanted to find out first whether this sort of intervention would be feasible, producing change in reading comprehension, and to assess whether there was any relation between discussions of meaning during training with posttraining comprehension improvement. Analyzing such conversation is time consuming and difficult, requiring the development of coding systems that capture metalinguistic discussion, but such analysis seems to be an important step in understanding processes that may be involved in comprehension development.

The first session of the ambiguity training involved introducing children to the idea that words can have multiple meanings, presenting homonyms, and asking children to generate homonyms. This session focused primarily on single words outside the context of a sentence. In Session 2, we introduced children to sentences with double meanings, such as "The man went hunting with a club." This move from homonyms to ambiguous phrases mirrors the progression used in the Test of Language Competence, which seeks to address "emerging metalinguistic abilities" (Wiig & Secord, 1989). In the third session, we introduced joking riddles with double meanings and helped the children to find the ambiguous words or phrases.

For a control condition, we developed three training sessions that also focused on metalinguistic aspects of language but that concerned sounds rather than meanings. Thus, in the different sessions we looked at homophones, tongue twisters and limericks, and jokes that played on the sounds of words. This allowed us to see whether it was the specific focus on meanings, rather than the general metalinguistic approach, that might foster comprehension change.

Method and Participants

Children were trained in small groups, and training sessions were audio-recorded, to enable us to analyze the extent to which ambiguity-trained children referred to alternative text meanings and then to assess whether there was a relation between sensitivity to meaning, as assessed by such comments, and comprehension improvement. We recruited 30 less skilled and 30 skilled comprehenders, using Yuill and Oakhill's (1991) criteria, according to which the two groups were similar in chronological age, single-word vocabulary, and reading accuracy but differed significantly in comprehension. On average, the poor comprehenders were 1 year behind their expected comprehension age, whereas the good comprehenders were 5 months above. All children were in school years 3 and 4 (ages 7–9) from schools in two large towns in southeast England. Twenty of the children in each comprehension group were given ambiguity training, in small groups of 4 to 6, and the remaining 10 in each group had small-group sound training.

The reading measure we used to assess pre- to posttraining change was the individually administered Neale Analysis of Reading Ability (Neale, 1989), in which children read aloud a series of short narrative texts that increase in difficulty. Any reading errors are corrected, and testing stops once the child makes a specified number of reading errors, so that reading comprehension scores are based on texts at the child's level of reading ability. The reading accuracy score is based on the number of words read correctly. The comprehension score is based on the number of questions answered correctly after the passage has been read. Questions are a fairly equal mix of literal and inferential. Within 2 weeks of completing the training, children were given an alternate form of the same test to assess posttraining comprehension.

Results

The results were somewhat complicated by variations in changes of reading accuracy score over training: Although the riddles training did produce improvements as expected for the poor comprehenders, the sound training also helped the good comprehenders, probably in part because this group was the poorest in reading accuracy at the outset: The training seems to have improved their word reading, which in turn helped them to answer more comprehension questions. An analysis of comprehension change from pre- to posttest, with changes in accuracy included as a covariate, showed a main effect of skill group, $F(1, 54) = 9.20$, $p < .005$, with poor comprehenders improving more than good ones, and a marginal interaction ($p < .06$) between skill group and training, as shown in Table 13.1, with striking gains made by poor comprehenders given riddles training. To understand better what processes might contribute to improvement, I now turn to analysis of the conversations children had during training.

There is little previous research in methods of analyzing metalinguistic discussion (but see Larkin, 2000) and apparently none that relates the quality

TABLE 13.1
Study 1: Mean change in Neale comprehension age as a function of skill and treatment group (scores adjusted for accuracy change)

Skill group	*Treatment*	*Change (months) in Comprehension Age*	
		M	*SD*
Less skilled	Ambiguity	17.38	2.58
	Sound	9.07	3.88
Skilled	Ambiguity	0.03	2.54
	Sound	4.59	3.78

and quantity of such discussion to outcome measures of comprehension. We therefore devised our own method of analyzing conversation for metalinguistic discussion and assessed the relation of such discussion to comprehension change. This allowed us to address two questions that seem to have been unexamined previously. First, to what extent are children of this age capable of making metalinguistic comments about alternative meanings? We can assess children's sensitivity to alternative meanings by counting the number of comments in the training sessions that mention multiple meanings. Second, is there a relation between metalinguistic comments and comprehension change? If comprehension changes are associated with growing awareness of multiple meanings, then one might expect increases across training sessions in comments indicating sensitivity to multiple meanings. By looking at discussions in the first and the third (final) training sessions we can gauge how much such awareness changes across time and relate it to the degree of change in comprehension. Thus, we expect there will be an association between the increase in comments on multiple meanings and the increase in comprehension score. The sample size clearly was very small in this exploratory study, so the data are presented purely descriptively and with appropriate caution.

We coded audio-recordings for the sessions for which we had complete data, the first and third sessions, with seven of the subgroups given meaning training: three groups consisting of less skilled comprehenders and four of skilled comprehenders.[2] Because sessions were only recorded on audio, and not on video, we could not reliably differentiate between children, so analyses are reported by group rather than by individual children. Making the group the unit of analysis does make some theoretical sense: Even if a particular child

[2]Since the time this pilot study was conducted, my colleagues and I have used mixed groupings or pairs of children for training, because we believe that this better supports children learning from each other, as I discuss later.

did not comment on a particular aspect of meaning, they were exposed to the comments of other children in the group and could learn vicariously. In the transcripts, we looked for any utterances that showed awareness of double meaning, and these comprised three types of utterance:

> Offering an alternative meaning once one meaning has been established, for example, "It could also mean water pipe."
> Offering two or more meanings within the same utterance, for example, "smoking pipe and water pipe."
> Spontaneously referring to double meaning using examples not mentioned in the training, for example, "What about hair? Head hair and animal hare."

We coded comments as "nonunderstanding" if children did not show explicit understanding, for example, repeating that words can have two meanings, but not providing any illustration, or offering a single meaning in two forms (e.g., "pipe and pipe cleaner" or "key and key ring").

Our question about whether children would make appropriate metalinguistic comments was given a strong positive answer: In both skilled and less skilled groups, almost one third of all utterances in Session 1, and around half in Session 3, fell into this category. Thus, with adult support and relevant material, children of this age can be engaged quite readily in metalinguistic discussion. There was also, though, quite a high level of nonunderstanding in both groups in Session 1, around 20% in Session 1 and 10% in Session 3.

Was comprehension change associated with increases in appropriate metalinguistic comments? Overall, the groups showed a significant increase from Sessions 1 to 3 in appropriate metalinguistic comments. Furthermore, the increase in such comments in this admittedly very small sample correlated with the increase in comprehension score, when changes in reading accuracy were partialed out, $r(4) = .80$, two-tailed, $p < .05$.

Discussion

The analysis of comments during training helps build a picture of what might be lacking in poor comprehension. Bearing in mind strong caution because of low sample size and the slightly unexpected pattern of results on the posttest for skilled comprehenders, the data are at least consistent with the idea that comprehension improvement may be fostered through comments highlighting the possibility of different meanings. This pattern of results lends support to the idea that comprehension failure may occur not because readers are incapable of making inferences but because they do not focus attention on text meaning during reading. Highlighting a focus on meaning through discussing ambiguous texts seemed to act as a stimulus for children to focus on meaning when they performed the posttraining comprehension test.

One informal observation that we made, which is not readily analyzable because of the small database, was that there seemed to be a clearer relation between high-level metalinguistic comments and posttest comprehension in Session 3 than there was for such comments and the comprehension pretest in Session 1. One possible explanation lies in a difference between the multiple meanings presented in Sessions 1 and 3 of the training. The first session introduced homonyms (e.g., *bank*, *fan*, *pipe*), whereas the third session used multiple meanings in joke contexts (e.g., the "spotted leopard" example mentioned earlier). There may be an important difference between simply producing multiple interpretations, unrelated to specific meanings in a text, and being able to select two different interpretations that are invited, or constrained, by the text. For example, a child may understand that the word *pipe* can mean an implement for smoking or a means to transport water, but he or she does not detect its meaning as a musical instrument, which might be the meaning required for a specific context: for example, a joking riddle that asks "Why is water musical?—Because it uses pipes."

In Session 1, children needed to retrieve different meanings for words out of context (e.g., "How many meanings can you think of for *pipe*"?), whereas in Session 3 we focused on different ways of interpreting word compounds within jokelike settings, where particular meanings were needed in order to "get the joke." For example, the joke "How do you make a sausage roll?—Push it down a hill," rests on an understanding that *sausage roll* can be a savory food item or the action of rolling a sausage down a slope. In the latter example, meanings are constrained by the intended meanings of a joke, in contrast to metalinguistic comments in Session 1, which involved producing multiple meanings outside a specific context. Although this difference between Sessions 1 and 3 was an accidental feature of the training, designed to introduce children in steps to the idea of multiple meanings, it suggests that a focus on meaning itself may be insufficient: Children require a focus on meaning and an understanding that meaning needs to be adapted to a particular context. Metalinguistic comments in Session 1 did not require context-sensitive interpretations and so may not be as important for comprehension skill. In contrast, such comments in Session 3 involved context-sensitive meanings and may relate more clearly to comprehension skill for children who may have previously lacked such skills.

Thus, multiple meanings may be used in a way that is or is not context sensitive. The importance of this distinction is shown in a study by Casteel (1997), who assessed age differences in understanding interpretive ambiguity by children aged 7 and 9 years. The younger children could offer multiple interpretations of ambiguous texts, but their second interpretations were less likely to be those intended by the authors. The older children, in contrast, constrained their reinterpretations to fit the text. Less skilled children could be characterized as developmentally less mature, showing a pattern like the younger children in Casteel's study. This supports the clinical observation that many poor comprehenders fail to focus on contextually appropriate meanings and do not constrain their inferences appropriately. The interpretation given is clearly post hoc and rests on a very limited database, but it is consistent with the idea that it is not just sensitivity to meaning that is important but sensitivity to meaning appropriately constrained by the context.

PEER DISCUSSION OF JOKING RIDDLES: EXPLORING MULTIPLE MEANINGS IN CONTEXT (STUDY 2)

The pilot study gave me and my colleagues some confidence that considering multiple meanings of words in context could play a role in increasing children's awareness of multiple meanings and, perhaps, help them develop a more reflective awareness of word and sentence meanings. In the training, presenting joking riddles containing ambiguous words or phrases was a highly motivating way of discussing different word meanings. To understand the joke, it is necessary to elucidate the different meanings by using disambiguating contexts, because a joke of this type rests on such context. As an example, consider this joke: "Does this restaurant serve fish?"—"Yes, what do you want to eat, Mr Fish?" The ambiguity rests in "serve fish." The expected or cued reading implied by the question is to assume that a person wishes to be served fish for a meal. The joke answer makes sense only if one reinterprets the question using an uncued sense of *serve*, to mean something like "Would you provide food to a customer who is a fish?" To explain the joke, one has to give the two contrasting contexts. As one excited child put it succinctly in explaining this joke, "I get it! 'Cause they serve fish on a plate and they serve fish to the fish."

We had also noted in the previous training that children sometimes seemed to benefit from the correct or incorrect answers other children gave and could be motivated to explain more explicitly if another child did not understand something. This might be explained in the same way as an unexpected finding by Yuill and Oakhill (1988) that undertaking comprehension exercises in groups led to some comprehension improvement, perhaps because children had to discuss and decide between the merits of different answers. We therefore felt that peer discussion of ambiguous material might foster comprehension change.

There were marked individual differences, as reported by the trainers, in our ambiguity training study in how readily children took to the task of explaining multiple meanings. There is therefore considerable scope for learning from peers through scaffolding (e.g., Wood & O'Malley, 1996). More specifically in the area of reading, Palincsar and colleagues (e.g., Palincsar & Brown, 1984; see also chap. 7, this volume) used Vygotsky's ideas about scaffolding to argue that children of different skill levels can support each other's understanding by discussing different interpretations of text. It appears that in unequal interactions, where one child understands more than the other, both children can benefit: One has the task explained by a peer who is not too far ahead of him or her in understanding, and the other is pushed by a partner's incomprehension to explicate further his or her own understanding.

Our studies showing the relation of joke understanding to comprehension led to the development of a piece of software, Joke City (Yuill & Bradwell, 1998), which presents pairs of children with jokes and encourages the children to explain them to each other. Players of the game have two main tasks: (a) to guess the answer to a given joke, and then, on being given the answer, (b) to pick out the ambiguity in the joke. To do this, they can use various verbal clues designed to focus their attention on the relevant ambiguity. After guessing at the ambiguity, children then see two pictures depicting the two different meanings embedded in the joke, and a verbal explanation. Completing these tasks involves the children in discussing the language in the jokes with each other. Pilot work showed that the software certainly engaged children, and involved them in sometimes quite prolonged discussion of meaning in context, but empirical studies were needed to assess the potential benefit of the software for comprehension, the nature of the conversations about language, and the ways in which this conversation might be related to comprehension change.

A pilot study of 14 pairs of 7- to 9-year-old children was conducted to develop a scheme suitable for coding children's conversations when engaged with the software. This study also suggested that pairing a good and a poor comprehender seemed to be the most effective method of generating discussion. This is in line with the Vygotskyan (1934/1986) idea of the *zone of proximal development*, whereby a higher level of thinking can be achieved under the guidance of a more capable peer. Although children no longer had adult guidance, peer pairings seemed to have real advantages. There were many occasions when children showed a clear metacognitive "a-ha" experience: The sudden insight that they got the joke, and the more informal atmosphere, led to many positive evaluations and much laughter. There were also examples of children on the edge of understanding; sometimes a remark by the other child in the pair came at just the right moment to trigger understanding,

TABLE 13.2
Study 2: Coding of utterances

Reading from the screen: any reading aloud of screen text, which included instructions (e.g., to click the mouse) and explanations of the jokes and double meanings. Children were required to do this as part of the task.

Control statements: any utterances involving the 'nuts and bolts' of the task, such as negotiating turn-taking, statements, questions or responses concerning performance of the task (e.g., requests to read), agreement or disagreement. Also included was any evaluation of the task, for example, 'This is fun.'

Metacognitive utterances about the state of children's knowledge and understanding, about what one knows or thinks, whether about the self, joint, or the other. These primarily concerned whether or not the joke was understood and comments about the correctness of a guess, for example, 'Aha I get it!', 'We tried that before', 'What do you think?'

Metalinguistic utterances, about the language of the jokes themselves (often prompted by questions posed in the software). These were subcategorized according to whether the utterance explained the cued meaning of the joke (as prompted in the question, to mislead the listener) the uncued meaning (the meaning one has to use to interpret the joke answer correctly),[1] or both meanings, for example, 'He wanted to draw [with a pencil] and draw the curtains.'

[1]For example, in the 'serve fish' joke mentioned in the text, the cued meaning of *serve* refers to the object first is served to eat, whereas the uncued meaning refers to the recipient to whom food is served.

although at other times a chance event or distraction prevented an apparent opportunity to get a correct answer. For a child who catches on more quickly, having a peer who does not understand might also avoid the demotivation that may have affected groups of good comprehenders in the pilot study described earlier in this chapter.

The coding scheme we devised for these conversations was, after several iterations, successful in encompassing all the talk and achieved a high level of reliability across two experienced coders (see Yuill & George, 2006). Reliability for each category was over 85% for blind double coding of four randomly selected half-hour sessions. Children's utterances fell into four very broad types, as shown in Table 13.2: (a) reading from the screen, (b) control statements, and (c) metacognitive and (d) metalinguistic utterances.

Given the feasibility of using Joke City, we performed a full training study to assess possible comprehension improvement after paired discussion of jokes, and to look further at the relation between metalinguistic talk and comprehension improvement.

Method and Participants.

We used the Neale Analysis of Reading Ability, as before, to select 12 pairs of children aged 7 to 9 (Years 3 and 4) from two schools in a large town. In each pair, we had one better and one poorer comprehender. The group available did not allow us to select poor comprehenders using the strict criteria of Yuill and Oakhill (1991), but the poorer comprehenders from each pair were significantly lower in comprehension skill than the better comprehenders in the pairs, and the better and poorer comprehenders across the pairs were not significantly different in their average age or reading accuracy. We also had 24 control children, matched for age, accuracy, and comprehension, who acted as controls, who just attended their normal literacy lessons.

Each pair of trained children participated in three sessions of about 20 minutes on Joke City, with increasingly higher (more difficult) levels of materials, over the course of 2 to 3 weeks. There were six jokes to disambiguate in each session. The trainer explained the software by guiding the children through the first joke at Level 1 and was present through the entire session but intervened primarily to keep the children on task, to ask for clarification of children's understanding when it was not clear (to facilitate later coding), and to give technical help if needed. Within 2 weeks on either side of this training, children were given one of two parallel forms of the Neale Analysis of Reading Ability. Each training session was videotaped, and the first and third session for each pair was transcribed and coded using the scheme described earlier.

Results

We found a clear overall training effect: Trained children improved in comprehension score (an average of 7 months on the age-converted scores). This was independent of changes in reading accuracy, which in this study were minimal. These results are encouraging, but there was marked variability between pairs in the extent of improvements and quality of discussion. We need to understand more about what processes might be conducive to improvement, so we can modify the intervention to focus more closely on understanding and encouraging those processes. For this reason, we next looked at the children's conversations during their training sessions. Because there were fairly small differences in the amount of talk in each pair, the results show a similar pattern regardless of whether proportions or absolute frequency of utterances are used.

The proportions of each type of talk were as follows: metacognitive, 19%; metalinguistic, 30%; control statements, 15%; and reading from the screen, 36%. A striking feature of the data is the extent to which children were capable of engaging in metacognitive and metalinguistic discussion; these two categories account for just under half of all the conversation. Given that children were working with a peer, this compares very favorably

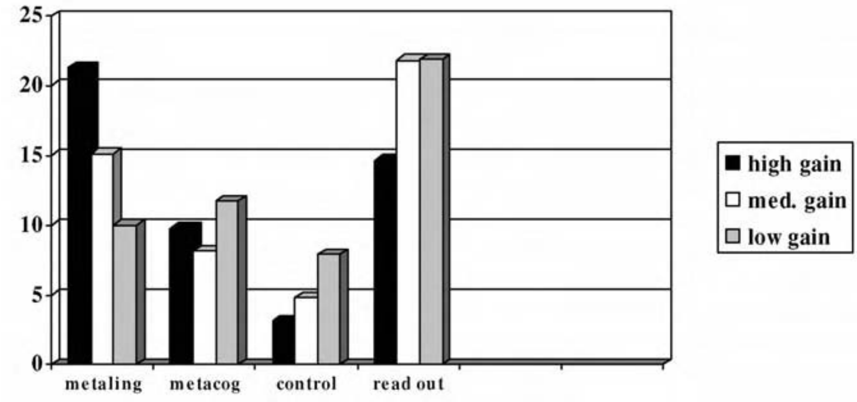


Figure 13.1. Study 2: Absolute frequencies of different talk types by improvement group. See text for definitions of categories.

with the amount of metalinguistic talk in our pilot study, in which an adult was constantly present to try to elicit such talk from children.

For some aspects of reading comprehension, it seems that *metacognitive skill*, awareness of one's own state of knowledge, is useful and worth training. However, for the current task our concern was to foster metalinguistic discussion of the ambiguities in meaning. We therefore expected that comprehension change would be related to the amount of metalinguistic talk rather than to the amount of metacognitive talk. This was indeed the case: The correlation between number of metalinguistic statements for each child and their increase in comprehension age (in months) between pre- and posttest is significant, both considering children individually, $r(22) = .54$, $p < .005$, and looking at the pair as a unit of analysis, $r(10) = .69$, $p < .01$. In contrast, the relation of comprehension improvement to the number of metacognitive comments showed no significant correlation with improvement, either for individuals or for pairs, $r(22) = -.05$, and $r(10) = -.33$, respectively.

Differences According to Extent of Improvement

Within the trained group, there were differences in the extent of improvement, and some children showed no improvement at all. Assessing the characteristics of more and less successful interactions is another useful way to help us understand processes of change. The 12 pairs fell into three fairly clear groups: 4 pairs who improved considerably (each child showing an increase of at least 5 months), 4 who showed mixed improvement (one child improving at least 4 months, and the other staying the same or dropping

TABLE 13.3
Study 2: Mean of Age and Pre test Scores and Changes in Accuracy and Comprehension Scores (in Months) for High, Medium, and Low Improvers and for Control Children

	Age		Accuracy			Comprehension		
			Pre test			Pre test		
Group	*M*	*SD*	*M*	*SD*	*Change*	*M*	*SD*	*Change*
High improvers (n = 8)	119.2	5.9	94.9	16.7	0.9	89.8	18.5	15.6
Medium improvers (n = 8)	117.8	5.9	90.2	11.4	–4.5	83.0	6.7	5.2
Low improvers (n = 8)	120.4	6.2	80.9	8.8	0.3	85.8	6.4	0.9
Control (n = 24)	111.2	5.2	83.4	11.4	–0.8	84.8	11.8	0.9

slightly) and 4 who showed little or no improvement over the pair (only –2 to +4 months' change each). The characteristics of these high, medium, and low improver subgroups are shown in Table 13.3; they were not significantly different in age or pretest comprehension, although the low improvers tended to be poorer decoders on the pretest reading accuracy measure, $F(2, 21) = 2.43$, $p = .11$, and a one-way analysis of variance of accuracy scores showed that the low improvers were significantly different from the high improvers in this respect (least significant difference, $p < .05$)

There were no significant differences in the amount of talk among these three groups, but there were clear differences in the nature of their talk, as shown in Figure 13.1. Metalinguistic talk was significantly higher in the high improvers, who made more than twice as many such comments as low improvers. Furthermore, high improvers showed a particularly big increase across the course of the training in comments that directly contrasted cued and uncued meanings in the jokes, supporting the hunch that these comments reflect a particularly clear understanding of meanings in relation to context. Here is an example from the discussion of the joke "Why don't leopards escape from the zoo?—Because they are always spotted." In the final utterance, Mark contrasts the cued and uncued meanings of "spotted" in a single statement:

> (On screen) What are the 2 different meanings? ("a leopard is always spotted")
> Mark: Spotted
> Andy: Well
> Mark: Well it can't escape 'cause it's got loads of spots.
> Andy: Spots
> Mark: And it's
> Andy: Spotted
> Mark: And it can always get spotted 'cause it's got spots.

There were no differences in other categories of talk, except that low improvers tended to spend more time reading from the screen. This probably occurred because their reading accuracy was poorer, so one child often had to reread or correct the misreadings of the other. Rereading also sometimes gave us the impression of being a way of filling in time when children were unsure of the answers.

GENERAL DISCUSSION AND IMPLICATIONS

The studies described in this chapter provide support from several sources for a link between sensitivity to meaning and comprehension and suggest the utility of discussing meaning in joking riddles as a way to foster general comprehension skills. The studies show the feasibility of some new paths of investigation into reading comprehension and raise new questions. The discussion brings out four points that arise from this work.

First, the evidence presented suggests that in addition to focusing on interventions concerned with text macrostructure, it is valuable to investigate comprehension at the sentence and phrase level, to help children make inferences that bring coherence to text. Second, it is clear that some children with poor comprehension just do not realize when they do not understand. One of the benefits of using joking riddles, as we did in these studies, is that they make meaning problematic, in a way that is familiar to children, so that they engage more naturally in the effort after meaning. There are different ways to engage children in this effort that have been used successfully in the past. For example, Carnine, Kameenui, and Coyle (1984) trained children to work out the meaning of novel words by showing them how to use context, and Yuill and Joscelyne (1988) wrote deliberately obscure texts to show children how to use words as "clues to meaning."

Our third point is that there are ways of examining processes of comprehension development, both in able and less able comprehenders. The method we used here is to look at change during intervention studies in relation to outcome. This method was possible because we trained children in pairs or small groups, and so we had access to their conversations about language, giving us an insight into their growing understanding in a way that would be very difficult to gain just by asking them. We know from metacognitive research that children (and adults) cannot always tell us explicitly about what or how they understand, and instructing them in how to use what we think of as "good" strategies does not necessarily work. As Tierney and Cunningham (1984) cautioned, "teaching children our theories about how they think in order to get them to think better seems to us to be fraught with danger" (p. 634). Engaging children with language using playful means such as joking riddles is an indirect way to help children develop sensitivity to meaning that may be more effective than a simple instruction to "pay attention to meaning." It is encouraging that children's conversations in the studies described in this chapter did relate to changes in outcome measures, but much more work needs to be done to see how these processes might contribute to children's developing metalinguistic awareness.

The final point is that reading research could usefully look outward to three other research areas—cognitive development, motivational studies, and human-centered technology—that can inform our understanding of the development of reading comprehension. We briefly address each of these in turn.

Studies in cognitive development tell us, for example, that children's understanding of the relation of language form to its meaning changes markedly in the early years, with periods of apparently rapid change in understanding homonymy around the age of 4 years (e.g., Doherty, 2004), and ambiguity around the age of 6 to 7 years (Robinson, Goelman, & Olson, 1983). Even so, Yuill, Kerawalla, Pearce, Luckin, and Harris (in press) reported that appreciation of homonymy and ambiguity continues to change beyond the age of 7 years and strongly predicted reading comprehension skill. They suggested that although relatively rapid changes may occur early in development, allowing children to understand in principle the relation between the form of text and its multiple possible interpretations, children still need to consolidate this understanding. One measure that predicts children's growing understanding of text interpretation is their reading-specific cognitive flexibility: the simultaneous coordination of phonological and semantic properties of words (Cartwright, 2002), and there is evidence that this flexibility is a good predictor of comprehension, independent of age, reading accuracy, and general verbal ability, also suggesting useful training methods.

Motivational aspects of comprehension were very clearly addressed in the studies reported here. There was little doubt that the children "visiting" Joke City saw themselves as playing with language: In analyzing the interactions, we found the need for a subcategory of metalinguistic talk that we called *metalinguistic play*, covering utterances that were not directly about the meanings of the ambiguity but were generally playful uses of language, such as making word associations from words in the joke, using the words in off-the-cuff songs, and playing with the sounds of the words. In particular, we found that boys seemed to see jokes as their domain of expertise and were happy to spend time discussing the language of jokes. In the final study in particular, the combination of computers and jokes was a potent one for the boys, and this is very valuable considering the often-expressed concerns about boys' reading skills and motivation. It should be noted, though, that jokes fit into a fairly narrow window of childhood, for many children at least: We found, consistent with the literature on joke understanding (e.g., Shultz,

1974) that children under about 7 did not understand the types of joke we used, and some children over about 10 to 11 years of age no longer found them as engaging, perhaps because they were seen as activities for younger children or because they were no longer as challenging.

Studies of motivational style distinguish broadly between two orientations to learning: (a) mastery, which focuses on understanding even at the expense of making errors, and (b) performance, with a stronger motivation to get the right answer. Joking riddles play to both of these: It is important socially to be seen to "get the joke," but at the same time one appreciates a joke only by understanding and resolving the apparent incongruity (Shultz, 1974), an ability that comes into play only after children have left the stage of just being able to recite a joke question and answer. Work in progress suggests that a mastery orientation is particularly useful when children are exploring the interpretive possibilities of jokes (Harris, 2006).

The third research area of potential use to literacy interventions is that of human-centered technology. This area should inform the increasing amount of work on automated interventions to deliver reading comprehension instruction. For example, if children are working together on a computerized reading task, how best can we represent information on the screen, and how can the system mediate children's interactions with each other to ensure genuine discussion and collaboration? Kerawalla et al. (in press) described a computerized task framework that encourages children to work simultaneously on their own representation of a literacy task while at the same time seeing their partner's representation, in a way that supports discussion and argumentation to reach a shared understanding of the task. This sort of collaborative interaction may lead to individual understanding that is superior to that developed by a child working alone on the task.

As the chapters in this volume show, there are many different aspects of comprehension and many different means of addressing comprehension difficulties. Discussion of word ambiguity appears to be one useful tool to foster comprehension progress, particularly for poor comprehenders who fail to make appropriate inferences. Embedding this in discussion of joking riddles is a highly motivating way to focus on meaning for children in this age group.

ACKNOWLEDGMENTS

I am very grateful to all the schools that took part; to Rebecca Lloyd-Lyons, Adam Galpin, and Karen Bain for their work on Study 1; to Nikki Khan (Study 2); to Pat George for help with coding and design; and for support from the Engineering and Physical Sciences Research Council, UK, Grant GR/R96538.

REFERENCES

Cain, K. (1999). Ways of reading: How knowledge and use of strategies are related to reading comprehension. *British Journal of Developmental Psychology*, *17*, 293–309.

Cain, K., & Oakhill, J. (1999). Inference-making ability and its relation to comprehension failure in young children. *Reading and Writing*, *11*, 489–503.

Carnine, D. L., Kameenui, E., & Coyle, G. (1984). Utilization of contextual information in determining the meaning of unfamiliar words. *Reading Research Quarterly*, *19*, 188–204.

Cartwright, K. (2002). Development and reading: The relation of reading-specific multiple classification skill to reading comprehension in elementary school children. *Journal of Educational Psychology*, *94*, 56–63.

Casteel, M. (1997). Resolving interpretive ambiguity in text: Children's generation of multiple interpretations. *Journal of Experimental Child Psychology*, *64*, 396–424.

Doherty, M. (2004). Children's difficulty in learning homonyms. *Journal of Child Language*, *31*, 203–214.

Francis, N. (1999). Self-correction patterns and metalinguistic awareness: A typology for studying text-processing strategies of proficient readers. *Journal of Research in Reading*, *22*, 304–310.

Gombert, J. E. (1992). *Metalinguistic development*. Chicago: University of Chicago Press.

Harris, A. (2006). *The role of learning goals in children's paired collaborative interactions*. Unpublished doctoral dissertation, University of Sussex, Brighton, UK.

Kerawalla, L., Pearce, D., Yuill, N., Luckin, R., & Harris, A. (in press). "I'm keeping those there, are you?" The role of a new user interface paradigm—Separate Control of Shared Space (SCOSS)—In the collaborative decision-making process. *Computers and Education*.

Larkin, S. (2000). *How can we discern metacognition in year one children from interactions between students and teachers?* Retrieved November 9, 2005, from http://www.TLRP.org/pub/acadpub/Larkin2000.pdf

Mahoney, D. L., & Mann, V. A. (1992). Using children's humor to clarify the relationship between linguistic awareness and early reading ability. *Cognition*, *45*, 163–186.

Neale, M. D. (1989). *The Neale Analysis of Reading Ability* (2nd ed., rev.). London: Macmillan Education.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*, 117–175.

Robinson, E. J., Goelman, H., & Olson, D. R. (1983). Children's understanding of the relation between expressions (what was said) and intentions (what was meant). *British Journal of Developmental Psychology*, *1*, 75–86.

Shultz, T. R. (1974). Development of the appreciation of riddles. *Child Development*, *45*, 100–105.

Tierney, R. J. & Cunningham, J. W. (1984). Research on teaching reading comprehension. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 609–654). New York: Longman,.

Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, Trans.) Cambridge, MA: MIT Press. Original work published 1934

Wiig, E. H., & Secord, W. (1989). *Test of Language Competence—Expanded edition*. City, State: Psychological Corporation.

Wood, D., & O'Malley, C. (1996) Collaborative learning between peers: An overview. *Educational Psychology in Practice*, *11*, 4–9.

Yuill, N. (1998). Reading and riddling: The role of riddle appreciation in understanding and improving poor text comprehension in children. *Cahiers de Psychologie Cognitive*, *17*, 313–342.

Yuill, N., & Bradwell, J. (1998). The laughing PC: How a software riddle package can help children's reading comprehension. In *Proceedings of the British Psychological Society* (Vol. 6, p. 119).

Yuill, N., & George, P. (2006). *Coding children's conversations for metalinguistic and metacognitive utterances*. Unpublished manuscript, University of Sussex, UK.

Yuill, N., & Joscelyne, T. (1988). Effects of organisational cues and strategies on good and poor comprehenders' story understanding. *Journal of Educational Psychology*, *80*, 152–158.

Yuill, N., Kerawalla, L., Pearce, D., Luckin, R. & Harris, A. (in press). Using technology to teach flexibility through peer discussion. In K. Cartwright (Ed.), *Flexibility in literacy processes and instructional practice*. New York: Guilford.

Yuill, N., & Oakhill, J. (1988). Effects of inference awareness training on poor reading comprehension. *Journal of Applied Cognitive Psychology*, *2*, 33–45.

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge, England: Cambridge University Press.

14

A Web-Based Tutoring System for the Structure Strategy: Theoretical Background, Design, and Findings

Bonnie J. F. Meyer
The Pennsylvania State University

Kay Wijekumar
The Pennsylvania State University Beaver

Instruction with the structure strategy has shown improvement on measures of reading comprehension for a wide variety of readers of all ages. Teaching this strategy primarily has involved traditional classroom instruction or one-on-one tutoring. We have created a Web-based intelligent tutoring system to present the strategy to middle school students. A total of 95 lessons (65 regular lessons and 30 parallel lessons for choice of text topic to boost motivation) containing 145 text selections were programmed into an intelligent tutoring system that was Web based, interactive, and multimedia based. The student and interaction models of the system included prior knowledge performance on the current task. The assessment tools included a parse tree based on the propositional analysis in the content structures and latent semantic indexing. Findings from pilot investigations in the development of the intelligent tutor are presented along with preliminary findings from a study run in two public schools.

Intelligent tutoring of the structure strategy (ITSS) is a Web-based tutoring program with an animated pedagogical agent tutor who teaches the structure strategy to middle school students. The goals of the ITSS are to help children learn to use the structure strategy to read and comprehend expository text. Our first functioning ITSS was aimed to match the learning needs and interests of students in fifth through seventh grades. In this chapter, we present the background theory and research that led to the development of the structure strategy and, ultimately, ITSS. In addition, implications of findings about structure strategy training are discussed in relation to reading theories. Next, aspects of the design of the Web-based system for ITSS are presented and followed by a description of aspects of the lessons. Finally, summaries of some of our recent findings are presented describing the use of ITSS with middle school students over 6 months.

THE STRUCTURE STRATEGY

The structure strategy teaches readers to identify the overall top-level structure of expository text (e.g., comparison, problem and solution, cause and effect, sequence, description, and listing) and to use that structure to organize their reading comprehension. *Comprehension* is defined as the synthesis of new ideas with existing memory structures using the framework of the structure strategy. Students are instructed in how to identify these commonly used structures and signaling (Meyer, 1975) used in text to explicitly cue these structures (e.g., “in contrast,” “on the other hand,” and “different,” for the comparison structure). They also learn a pattern for writing the main idea using each of the different text structures. For example, for the comparison structure, the pattern is: _______ and _______ (two or more ideas) were compared on ______, ________, and _______ (*X* number of issues compared). Next, students learn to use their main idea and selected structure to organize their reading comprehension and recall.

Learning the structure strategy is not only about structure (Duke & Pearson, 2002) but also about understanding the logical structure connecting ideas in a content domain. The content is important, too, and the purpose of learning the strategy is to increase understanding and comprehension of such content. For example, in the current ITSS there are eight articles of varying lengths and complexity related to the Pony Express (at least one using each of the five major structures), ranging from an advertisement calling for riders, to contrasting Wild Bill Hickok and Buffalo Bill Cody, and on to the effects of the transcontinental telegraph on the Pony Express. Readers also learn how to integrate ideas between different passages with the five basic structures.

They also learn how an author's purpose to inform or persuade a reader relates to the text structures the author uses.

Pressley and McCormick (1995) explained that the structure strategy is important in improving the reading comprehension skills of late elementary school students because it enhances students' ability to "[use and analyze] text structure to abstract main ideas" (p. 480). Meyer, Brandt, and Bluth (1980) found only 48% of a sample of ninth-grade students systematically used a plan, when asked to recall two texts (problem and solution and comparison top-level structures).

For example, these students organized their recalls from a problem and solution text in two parts. The first part contained information about the problem described in the text, and the second part described a solution that could potentially get rid of a cause of the problem and reduce or eliminate the problem. The remainder of the students simply listed some things they remembered with no attempt to interrelate the ideas into a coherent whole. In this study, most good readers (as measured by standardized reading achievement tests) recalled the top-level structure of the passage and used it to organize their recall, whereas most poor readers did not. However, students can learn to identify and use the top-level structure of text by learning the structure strategy (e.g., Bartlett, 1978; Englert & Hiebert, 1984; Meyer, Young, & Bartlett, 1989, Meyer et al., 2002).

To better explicate these findings about the structure strategy, we will look at the performance of a fifth-grade student who participated in the Meyer et al. (2002) study and made outstanding gains in reading comprehension. Before training, she scored at the 28th percentile on a standardized reading comprehension test and did not use the structure strategy, whereas after training she showed mastery of the strategy and scored at the 68th percentile on the reading test. Her pretest and immediate posttest recall of problem and solution texts are reproduced on the next page in this chapter; she recalled 9 ideas on the pretest and 88 ideas on the posttest. Most (70%) of the students in a control group without training did not organize their ideas with a problem part and a solution part; most completely missed any suggested solutions (how to prevent oil spills from supertankers or how to prevent deadly stings from killer bees explained in the two texts counterbalanced over pretest and posttest in the study). Their recalls were similar to that of this participant of the pretest before training with the structure strategy (see pretest recall from the supertanker text in the following paragraph). However, after training, this fifth-grade student has both a problem part, signaled with "problem," and a solution part, signaled with "one way." Although not as clear as it could be, the last three sentences are an attempt to recall the solution part of the passage.

Pretest: This passage is about oil spills. The oil spills on the ocean and poisens them. When the oil spills it kills animals too and, poisens them. I can only remember something about 3 football fields.

Posttest: The problem is prevention of killer bees. Bees make honey 150 pound per year. They reproduce quickly in warmer climates. They don't live under 59 decreas. Some of them escaped from Africa and came to S.A. Brazil. If their nest is desterbed they will sting. One man was riding his horse in Brazil and the bees came up and started stinging him and his horse. He fell from his horse and survived, but his horse died because of all the stings from the bees. Bees can not see red, that's why bee keepers wear red when working with bees. A lot of bee strikes can kill a person. Mostly they live up to North Carolina. Dust can calm the bees. One way scientists teach the people of Brazil is don't desterb their nests and run from killer bees. Scientists can't stop all the killer bees.

Instruction about text structures has yielded positive effects for reading comprehension with children and young and older adults (Armbruster, Anderson, & Ostertag, 1987; Bartlett, 1978; Carrell, 1985; Cook & Mayer, 1988; Englert & Hiebert, 1984; Gordon, 1990; Meyer, 1999; Meyer & Poon, 2001; Meyer et al., 2002, Meyer, Talbot, Poon, & Johnson, 2001; Meyer et al., 1989; Paris, Cross, & Lipson, 1984; Polley, 1994; Raphael & Kirschner, 1985; Richgels, McGee, Lomax, & Sheard, 1987; Samuels et al., 1988; Slater, Graves, & Piche, 1985; Taylor & Beach, 1984; Weisberg & Balajthy, 1989; Williams, Hall, & Lauer, 2004; Williams et al., 2005). Gains in reading comprehension have been reported for children as young as second grade (Williams et al., 2005) and for older adults in their 80s (Meyer & Poon, 2001).

In many respects, the structure strategy is similar to Vitale and Romance's (see chap. 2, this volume) propositional concept mapping; both focus on the hierarchical, logical structure of a text. The goals of the two approaches vary more on emphasis than in essence. Vitale and Romance's emphasis is on building knowledge of important ideas in science with the mapping of interrelationships seen as a tool in service of this goal, whereas our emphasis focuses on using text structure strategically with secondary goals of learning content in various domains.

Theodorou (2006) found that teaching the problem and solution structure in one training session with only medical content to college students resulted in transfer of the problem and solution structure when reading other medical texts but not texts in other domains, such as the environment. An important component of ITSS is that students read many different types of expository and persuasive texts about different topics that use the same overall top-level structure; the current ITSS has 65 lessons with 30 parallel lessons and 145 texts (some lessons have more than 1 text) ranging from 13 to 814 words with an average lexile grade equivalent of 5.43 (The Lexile Framework for Reading, 2005). Topics of the texts focus on science (34%), social studies and history (28%), animals (23%), sports (9%), and foods or recipes (6%). The diversity of content and style of texts in ITSS and the explicit instruction in the structure strategy help students transfer the strategy to text outside of the computer laboratory. With an early, nonautomated, Web version of ITSS (Meyer et al. 2002), the superiority in amount of information remembered from text by the structure strategy group over the control group was clearly evident in a large-group testing session conducted in a cafeteria (quite a different setting from instruction in the computer laboratory) 2½ months after training. A student scoring at the mean for the group receiving tutoring with the structure strategy training had a total recall score equal to a reader in the control group who scored at the 81st percentile on the delayed posttest (effect size = 0.92).

THEORETICAL AND RESEARCH BASE FOR ITSS

In this section, we discuss the theoretical and research base for the ITSS. The beginnings of this line of research started in the early 1970s. Issues studied then, plus those further developed in recent years, relate to contemporary issues in discourse theory, particularly models positing coherence building.

Interaction Among Text, Reader, Strategy, and Task Variables

We view performance in reading comprehension as an interaction of the text, reader, strategy, and task variables (see Figure 14.1). Details on how text and reader variables interact were explored by Meyer (2003), and details about interactions among reader, text, and task variables were explicated by Meyer and Rice (1989). The text variables of primary interest for ITSS are text structure, topic, and signaling (Meyer, 1975). For example, sometimes the author does not provide explicit signaling of the structure organizing the main ideas of the text, and therefore the reader needs to identify what the structure should be. “Research shows that signaling the organization of text or signaling main ideas yields superior recall of main ideas (e.g., Dee-Lucas & DiVesta, 1980; Loman & Mayer, 1983; Lorch & Lorch, 1985, 1995, 1996; Meyer et al. 1980; Meyer & Rice, 1989)” (Meyer, 2003, p. 214). Therefore, the organization of text affects how the readers organize their understanding and memory. ITSS uses a variety of texts that are explicitly signaled for the middle school students to learn how to use the author’s organization to improve their understanding of the text. There is some fading of explicit

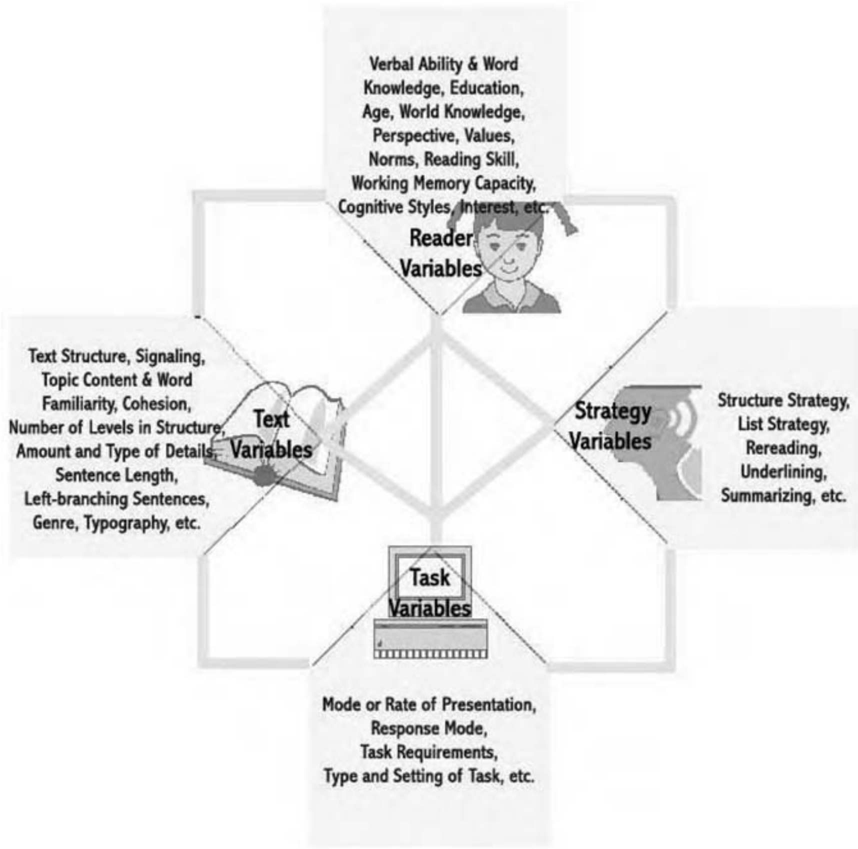


Figure 14.1 Interaction of text, reader, strategy, and task variables.

signaling as a student progresses to later lessons so that some practice is provided for determining text structure without signaling.

The reader variables of primary interest for ITSS are age, interests, standardized measures of reading comprehension and vocabulary, pretest performance measuring strategy use, progress in proceeding ITSS lessons, motivation, and self-efficacy. The strategy variable of interest is the structure strategy.

Some of the text, reader, task, and strategy variables that affect reading comprehension are presented in Figure 14.1. For an example, we will examine the early Meyer et al. (1980) study to show an interaction among a strategy variable, a text variable, and a reader variable. Meyer et al. (1980) proposed two dominant reader strategies for ninth-grade readers: (a) the structure strategy and (b) the list strategy. Readers using the list strategy encode text as a list of facts to be learned. Some readers who predominantly use a list strategy switch their strategy to the structure strategy when signals highlight the superordinate organization of the text. Both Meyer and colleagues (Meyer et al., 1980; Meyer & Rice, 1982, 1989) and Mayer and colleagues (Loman & Mayer, 1983) have suggested that signals cause readers to switch their strategies for encoding text information. Signals appear to influence the processes of selecting and organizing information in text by readers. Meyer et al. (1980) found that signaling enabled underachieving ninth-grade students, whose reading comprehension test scores were substantially lower than their vocabulary scores, to switch their strategy to the structure strategy from the list strategy when reading a text about supertankers. However, they found no effect of signaling for ninth-graders identified by teachers and standardized tests as exceptionally high in reading comprehension. Most of these skilled readers used the same organization as the author to organize their recall and maintained high levels of recall regardless of the presence of signaling. Meyer et al. (1980) also reported that ninth-grade students with both poor comprehension and vocabulary scores (perhaps reflecting poor word attack skills) exhibited the list strategy regardless of signaling. Therefore, the research conducted with ITSS used students' reading skill level measured scores on standardized reading comprehension tests and teacher appraisals to stratify and randomly assign students to research conditions.

The task variables of interest to our work with ITSS relate to computer presentation, mode and type of responses, and setting of delivery (i.e., required school learning tasks, electives in school and after-school settings, etc.). Task variables such as computer presentation versus printed materials have been shown to interact with the reader variable of age:

Meyer and Poon (1997) found that presenting information on computers enabled young adults to learn more efficiently than reading from text printed on paper, while presenting older adults with information on computers impeded their performance that was more efficient when reading from more traditional printed page (Meyer, 2003, p. 204).

The students participating in the current research study were in Grades 5 through 7 (volunteers with parental consent and student assent) and ITSS was a completely Web-based presentation of the task. The fifth- and seventh-grade students used ITSS during the school day. Students had to opt out of study hall, an attractive source of social engagement for seventh-grade students, to participate. The ITSS program was incorporated into the regular curriculum during the school day in some schools and during the after-school programs in other schools in Pennsylvania. Also in 2006 we individualized the sequencing and number of lessons, amount of audio help with reading texts, and difficulty level of text for students with different needs in the hope of maximizing learning of the structure strategy and improvement in reading comprehension.

Theoretical Background of the Structure Strategy

Cognitive research on reading comprehension has historically seen three major thrusts, but all continue as active areas of research today (see van den Broek, Young, Tzeng, & Linderholm, 1999). In the 1970s, the focus of the research was memory representations and what readers remembered after reading (e.g., Crothers, 1972; Kintsch & van Dijk, 1978; Mandler & Johnson, 1977, Meyer, 1971, 1975; Meyer & McConkie, 1973). For example, the 1970s focus was the logical structure of a text specified with some ideas high in a hierarchy of importance with many ideas connected to them, whereas other ideas were peripheral and low in the structure with few connections to other ideas (Meyer & McConkie, 1973). Effects of this logical structure were seen in the kinds of ideas that were remembered (more high-level information than low-level information—the levels effect); intraindividual stability in ideas recalled; and clustering of recall, suggesting that the structure was related to constructed cognitive structures (if a particular idea were recalled, the idea above it in the logical structure was recalled 70% of the time, rather than the average 23% recall of all ideas). The logical structure accounted for much of the variance that might ordinarily be attributed to other variables: serial position effects and rated importance.

Combining the approach of Meyer and McConkie (1973) with the linguistics work of Fillmore (1968) and Grimes (1975), Meyer (1975) specified the organization of text showing both local and global coherence in a content structure. The *content structure* is a propositional structure that specifies all the concepts in the text and how they are interrelated. Ideas high in the structure correspond to main ideas, whereas specific, minor details are found low in the structure. Meyer (1975) found that the pattern and type of relationships high in the content structure affect reading comprehension and recall, whereas these types and patterns have little effect on recall when low in the structure. Because of this finding, which was revealed by an examination of the correlations of the frequency of recall of propositions high and low in the structure while varying content but keeping the structure the same, Meyer started to focus her research on these different structures that organize the main ideas at the top levels of the content structures (e.g., Meyer et al., 1980).

Meyer's content structure has similarities to Kintsch's (1998) textbase and similarities to his situation model because causal, time, comparative, and other relationships are specified in the content structure. According to Graesser's outline of theoretical levels of text representation (see chap. 1, this volume), the content structure provides both propositional textbase information and rhetorical structure/text genre information.

A new focus in reading comprehension research during the 1980s was online processes in reading using gaze durations with eye-tracking techniques, reading times for words and parts of sentences, lexical-decision latencies on tested words, recognition latencies, probing techniques, and so on. Models focused on cognitive processes that take place online, such as what ideas are activated, which inferences are made, and how processing relates to limited working memory capacity (e.g., Graesser, Singer, & Trabasso, 1994; Kintsch, 1988; McKoon & Ratcliff, 1992). The construction portion of Kintsch's (1988, 2004) construction-integration (CI) model is an example of bottom-up processing models used to understand online processes in reading. Instead of trying to find only one correct meaning of a sentence, the CI model constructs several possible meanings in parallel and later sorts out the best meaning of the sentence. The sorting is accomplished by an integration process that inhibits constructions that do not fit well with the emerging context of the text and strengthens the constructions that do. Activation is spread around an incoherent propositional network with contradictory assumptions that is then cleaned up by the integration process, which results from the stabilizing of spreading activation, eventually settling on those nodes in the structure that hang together. The outliers and isolated nodes become deactivated in this bottom-up integration process. The model was based in part on the data of Till, Mross, and Kintsch (1988) using the lexical decision technique. Participants read sentences such as "A beautiful sight in downtown Denver is the mint." Immediately after reading the sentence, response times were short for both *money* and *tea*, exemplifying the construction portion of the CI model. This is a passive, implicit, low priming of long-term memory called *resonance* by some theorists (e.g., Myers & O'Brien, 1998). Delaying the lexical decision task by 350 milliseconds yielded short times for *money* but not for *mint*; the differential priming after the delay provides evidence for the integration portion of the model. However, little research has focused on how the more top-down, inference-generation process of integration actually works (Long & Lea, 2005). Our findings and those of others (chap. 8, this volume) related to structure strategy instruction should influence theorizing related to more top-down integration processes.

The most recent focus of cognitive research on reading comprehension is integration between the former two approaches of memory representation and comprehension processes and how the two relate. An example of this approach is van den Broek et al.'s computational model, called the *landscape model* of reading (e.g., van den Broek at al., 1999; van de Broek, Rapp, & Kendeou, 2005). The model incorporates both bottom-up activation (automatic spread of activation for associated concepts) and top-down processes aimed at seeking coherence (strategic, goal-directed searches for meaning; e.g., Graesser et al., 1994). Van den Broek et al. (2005) conducted three simulations of reading expository text to see whether both cohort activation and a top-down focus on major causal and referential connections among ideas, called *coherence*, were required to predict what readers recall. With both the cohort activation and coherence retrieval processes, accumulated node strength in the model correlated strongly with ideas recalled ($r = .70$); when cohort activation was removed, predictive power decreased ($r = .60$). When coherence-based processes were removed and cohort activation remained, the simulation's predictions fell further ($r = .50$). This last finding indicates the importance of coherence-based processing.

Reading comprehension relies on both the strategic and goal-direction search for meaning and overall logical structure and the more autonomous and passive processes. The structure strategy is one such technique that focuses on seeking coherence among text ideas. Readers use their knowledge of text structures to build coherent memory representations, and these structures or relationships are part of their cognitive representation. Signaling words for these structures guide readers toward coherent text representations with their key role in selection and encoding (e.g., Lorch & Lorch, 1995; Meyer & Poon, 2001).

There is online evidence by means of reading times and sentence verification tasks as well as recall data pointing to the psychological reality of these structures (Meyer & Freedle, 1984; Sanders & Noordman, 2000). Text structures not only describe text but also are cognitive entities in coherence representations of good readers. These structures were even explicated by Aristotle (1960) as ways to arrange the structure of a discourse or to guide invention of ideas for writing. In ITSS, after a new structure is modeled with the structure strategy and thoroughly practiced, it is used in a writing activity wherein students are given a structure, some signaling words, and a group of content words to select from as they create and organize a text with the specified structure and appropriately signal it.

Tables 14.1 through 14.3 display the order, number, and type of the 65 ITSS lessons (along with 30 parallel lessons for alternative choice of topics to increase motivation during practice lessons). For each structure, the intelligent tutor (IT) models the strategy and then gives help at each stage of identifying signaling, naming the structure, writing a main idea with the text on the screen, and recalling the text with the text off the screen. Next are approximately 5

TABLE 14.1
Order, Number, and Type of Lessons in Intelligent Tutoring the Structure Strategy for Comparison and Problem-Solution Structures

Order of lesson	*Activities engaged in by student*	*Example texts in standard lessons*	*Readability lexile GE*[a]	*Parallel text for topics choice lessons*
		Comparison structure, lessons 1–12		
1	SI,M	Washington and Lincoln	2-3	
2	Modeling	Washington and Lincoln	2-3	
3	SI,M,R	Elephants	2-3	Crocodilians
4	SI,M,R	Whales	4	Bears
5	SI,M,R	Babe Ruth, Hank Aaron, Barry Bonds	6-7	Sammy Sosa, Mark McGwire, Barry Bonds
6	ST,SI,M,R	Two different dogs	4	Two different parrots
7	ST,SI,M,R	Olympic women	4-6	Olympic men
8	M	Flying squirrels Caffeine study Lake vs. city	4 5-6 7-12	Flying squirrels Children with allergies Lake vs. city
9	Paper test	Colonial families	4-5	
10	Correct work of other students on prior text, SI, M, & R for Bats on Paper Test	Bats—leaf nosed and black	4-5	
11	SI,ST,M,R	Dogs and cats	2-3	Chinchillas and potbellied pigs
12	Titles to show structure, create title, MC	Squirrels Lake vs. city Washington and Lincoln	4 7-12 2-3	
		Problem & Solution structure, lessons 13–20		
13	Modeling problem & solution, SI, M	Whales	6-9	
14	R	Whales	6-9	

(*Continued*)

TABLE 14.1 (*Continued*)

Order of lesson	*Activities engaged in by student*	*Example texts in standard lessons*	*Read-ability lexile GE*[a]	*Parallel text for topics choice lessons*
15	SI,ST,M,R	Dog troubles—Porkchop	4-5	Pig Troubles - Hamlet
16	Question answer type of prob/sol; ST,SI,M,R	Washington's smile?	3	Taft sleeping?
17	SI,ST,M,R	Slimming plump dog	4-6	Slimming plump cat
18	ST,SI,M,R	Rabies in cats	5-6	Rabies in ferrets
19	ST,SI,M,R	Dust mites	10-12	
20	SI,M,R	Heartworm in dogs	5-6	Heartworm in cats
		Reviewing and integrating 2 learned structures 21–24		
21	Writing with 2 structures	Writing with comparison & problem & solution structures		
22	Author's purpose, MC inform or persuade, D, SI, M	Presidents Lake and city Reynolds Wrap advertisement Ben Franklin's lightning rod	2-3 7-12 1-2 4-5	
23	SI,ST,M, MC, inform or persuade,	Olsen twins Bill of Rights Madison bill of rights Popcorn Snow shovel	4-6 6-9 10-12 6-7 6-8	
24	Problem & solution structure with comparison of contrasting solutions, SI, D, M,R	Combining structures in a passage about rabies in humans	6-9	

Note. [a] Activities engaged in by students: SI = Click on Signaling word, M = Write Main Idea, R = Write Full Recall; ST = Write Name of the Structure, MC = Multiple- Choice Questions. D = Filling in or Examining Diagrams of Structure;
[b] Scores or range of scores indicate readability grade equivalents for the passage based on lexile scores: see The Lexile Framework for Reading. "Lexile analyzer." Retrieved March 13, 2005 from http://www.lexile.com/DesktopDefault.aspx?view=re&tabindex=2&tabid=31&tabpageid=358.

TABLE 14.2
Order, Number, and Type of Lessons in Intelligent Tutoring the Structure Strategy for Cause and Effect and Sequence Structures With Review of Other Structures

Order of lesson	*Activities engaged in by student*	*Example texts in standard lessons*	*Read-ability lexile GE*[a]	*Parallel text for topics choice lessons*
		Cause & effect structure, lessons 25–34		
25	Modeling cause & Effect, SI,R	Hailstones	5-6	
26	Modeling D,M,R	Whales Heat causes changes in rate of aging eye	 10-12	 Whales Heat causes changes in skin cancer growth
27	SI,MC, inform or persuade, M,R	Ice cream	3-4	
28	SI,ST,D, M,R,	Johnstown flood	6-8	
29	SI,ST,M,R	Milkshake recipe Cause of dog cage chaos	5-6 9-12	Sticky buns recipe Cause of bird cage chaos
30	SI,ST,D, M,R	Chicken hawks	5-6	
31	SI,ST,M,R	Reason for raining frogs	5-6	Reasons for raining fish
32	SI,ST,D, MI,R	Connected telegraph wires	4-6	
33	M,R	Hailstone effects in India	4-5	Hailstone effects China
34	SI, ST,D, M, Paper test for R	Pony Express	5-6	
		Reviewing and integrating 3 learned structures, lessons 35–40		
35	Paper test	Supertankers	6-9	
36	SI,ST,D Review prior structures with multiple texts	Sewage Wildfires White snowflakes Litterbox odors Living room 1928 vs. today George Washington Duck dead from botulism Grand Coulee Dam vs. Hoover Dam	6-8 7-12 4-6 3 5-6 4-6 6-8 11-12	

(*Continued*)

TABLE 14.2 (*Continued*)

Order of lesson	*Activities engaged in by student*	*Example texts in standard lessons*	*Read-ability iexile GE*[a]	*Parallel text for topics choice lessons*
37	Writing: 3 Structures	Writing modeled with cause & effect: Students practice it and also 2 prior structures		
38	SI,ST,M,R	First part of Basic training text	6-8	
39	M, D	Pony Express	5-6	
40	Paper test	Killer bees	5-6	
		Sequence structure, Lessons 41–49		
41	SI,ST,M modeled R	William Cody	5-6	
42	SI,ST,M,R	George Washington	6-8	Abraham Lincoln
43	SI,ST,M,R	Baby teeth timeline	4-6	Kitty teeth timeline
44	SI,ST,M,R	History of a state: Pennsylvania	5-6	History of a state: New York
45	SI,ST,M,R	History of the bald eagle	5-6	History of spotted owl
46	SI,ST,M,R	Biography Benjamin Franklin	6-8	Biography Clara Barton
47	SI,ST,M,R	Biography William Penn	4	Frederick Douglass
48	SI,ST,M,R	Biography Michelle Kwan	4-5	Andre Assai
49	Paper test	Biography Wild Bill Hickok	4-6	
		Reviewing & integrating 4 learned structures lessons 50–52		
50	Writing: 4 structures	Sequence writing modeled: Students practice with sequence & 3 prior structures		
51	SI,ST, review prior structures: compare, probllem & solution, cause & effect, sequence	Pollution Six steps to CPR Snowflakes Thoe's Tasty Chunky Tuna Ben' Franklin's lightning rod Innocent gesture Ice cream Happy heart Kemp computer solution Clouds	2 3 2-3 5-6 5-6 6-8 4-5 10-12 4-5 6-8	

(*Continued*)

TABLE 14.2 (*Continued*)

Order of lesson	*Activities engaged in by student*	*Example texts in standard lessons*	*Read-ability lexile GE*[a]	*Parallel text for topics choice lessons*
52	SI,ST,M,R	Combining Structures: Modeled with How-to purchase and Practiced with How-to study in class	3-4	

Note. SI = click on signalling word; R = write full recall; D = filling in or examining diagrams of structure; M = write main idea; MC = multiple-choice questions; ST = write name of the structure. [a]scores or ranges of scores indicate readability grade equivalents (GEs) for the passage based on lexile scores: See The Lexile Framework for Reading (2005).

practice lessons on the structure that decrease in scaffolding provided by IT. Then the students are involved in a formative evaluation recall task; some instruction is provided online as well as the targeted text for recall, but recall is requested on paper (a clipboard with paper) after the text is removed from the screen. Next, once two structures have been presented and lessons have been complete in which the structure is integrated with previously studied structures, comes the writing task. Then the lessons focus on more complex texts and combining different structures to show how they work together to convey an author's main points.

The following is the text for Lesson 5, which was designed for students to practice the comparison structure by comparing two different dogs:

Two Very Different Dogs

> The Jello family has had two very different dogs. Winky is the Jack Russell Terrier that the family has now. He is a one-person dog. Winky likes Bonnie best in the family. Winky growls if a family member tries to pick him up when he is sleeping on Bonnie's lap.
>
> In contrast, Dakota is a mixed-breed dog that the Jello family had five years ago. Unlike Winky, Dakota never growled and liked everyone in the family the same.

The parallel Lesson 5, designed for providing students with a choice of topics (dogs or parrots), compared two parrots:

Two Very Different Parrots

> The Jello family has had two very different parrots. Zeus is the African Grey Parrot that the family has now. He is a one-person parrot. Zeus likes Christina best in the family. Zeus bites if a family member tries to pick him up when he is perched on Christina's hand.

TABLE 14.3
Order, Number, and Type of Lessons in Intelligent Tutoring the Structure Strategy for Description and Combined Structures

Order of lesson	*Activities engaged in by student*	*Example texts in standard lessons*	*Read-ability lexile GE*[a]	*Parallel text for topics choice lessons*
		Description Structure Lessons 53–62		
53	Modeling of Description SI	Swans	4-6	
54	SI,ST,M,R	Korean cuisine	5-6	Japanese cuisine
55	SI,ST,M,R	Six Flags amusement parks	5-6	Disney World amusement park
56	SI,ST,M,R	grizzly bears	5-6	Panda bears
57	SI,ST,M,R, MC inform or persuade	Advertisement: Eveready battery	2-3	
58	ST, D,M, R	Diamond Walnuts	9-12	Popsicle sticks
59	SI,ST,M,R	Meet a kid poet and painter	3-4	Meet a kid runner and quarterback
60	ST,M,R	Spending time at the barn	4-5	Spending time around the truck
61	SI,ST,M,R	Rattlesnakes	4-5	Wild turkeys
62	SI,ST,M,R, D on paper	Tornados with diagrams and Hurricanes on paper	6-9	
		Reviewing & integrating 5 structures, lessons 63–65		
63	Writing: 5 structures	Modeling Description Writing with "My favorite candy store" and students practicing writing with the description structure and other structures	4-5	
64	SI,ST,D, Review of all 5 structures with multiple texts	Hickok	12+	
		Hickok and Cody	5-6	
		Field hockey	5-6	
		Tasters	4	
		Football equipment	6-9	
		Steamboats	4-6	
		Wanted Pony Express riders	5-6	
		Aesop tree	6-9	
		Keeping light on Pony Express	5-6	
		Thanksgiving	5-6	
		History of communications	4-5	
		Paterno vs. Bowden	2-3	
		Broccoli	2-3	

(*Continued*)

TABLE 14.3 (*Continued*)

Order of lesson	*Activities engaged in by student*	*Example texts in standard lessons*	*Read-ability lexile GE*[a]	*Parallel text for topics choice lessons*
65	Integrating Structures, SI,ST,MC, M,D in Multi session final lesson	Integrating 5 Structures Using a Long Magazine Article: Basic Training (about dog training; 814-word text)	4-5	

Note. SI = click on signalling word; ST = write name of the structure; M = write main idea; R = write full recall; MC = multiple-choice questions; D = filling in or examining diagrams of structure. [a]scores or ranges of scores indicate readability grade equivalents (GEs) for the passage based on lexile scores: See The Lexile Framework for Reading (2005).

In contrast, Quinday is a Nanday Conure Parrot that the Jello family had 10 years ago. Unlike Zeus, Quinday always bit and disliked everyone in the family the same.

Next we describe the ITSS interface and present the research supporting the design choices for the Web-based system.

DESIGN OF THE WEB-BASED ITSS

ITSS brings together multiple research themes on motivation, multimedia learning, metacognition, and memory structures to create a cycle of interactions and implemented using a .NET and Flash Web-based interface shown in Figure 14.2. These themes were synthesized from current research on intelligent tutoring technologies (Anderson, Corbett, Koedinger, & Pelletier, 1995; Koedinger, 2001; Lajoie, 2000; chap. 16, this volume; McNamara, 2004; McNamara, Levinstein, & Boonthum, 2004) and combined with ideas from multimedia learning research (Mayer, 2001). ITSS combines the intelligent tutoring technologies with multimedia delivery schemes to model strategy use, allows students to practice using the strategy, assesses the student responses (and creates a student model), and provides immediate feedback (based on the student and interaction models). The intelligence in the ITSS comes from the student and interaction models that use database-driven rules and latent semantic indexing (LSI) to continuously update and improve the interactions. Next, we briefly describe some of the extant research that contributed to the design of ITSS.

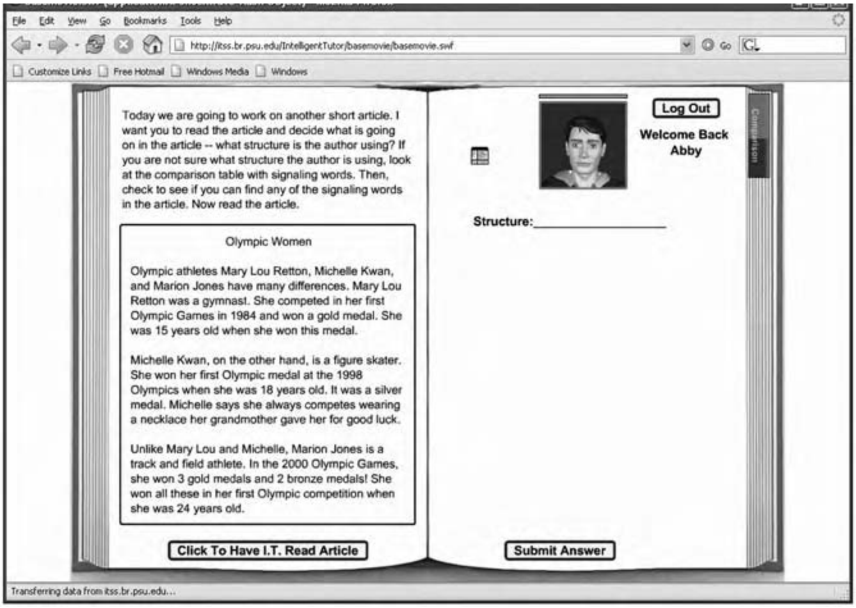


Figure 14.2 Intelligent tutoring of the structure strategy Flash interface. I.T. = intelligent tutor.

Interaction Cycle Step 1: Modeling

Modeling is a fundamental unit of multimedia learning environments (Mayer, 2001) and computerized tutoring environments (Anderson et al., 1995). The purpose of modeling is to show the learners how experts would solve the problems, the steps that they use, the critical focal points in the problems, pointing out pitfalls, and allowing the students to observe and learn the problem-solving process. In the case of reading comprehension, the purpose of modeling is to show the students how experts would read prose, organize their memory structures, and construct their recall.

Applying the structure strategy to reading and remembering more information requires the user to do the following three things: (a) understand the five text structures as well as the signaling words that aid categorization of texts into these structures; (b) construct a main idea for the text using the structure; and (c) recall the information using the structure, main idea, and appropriate signaling words. In previous research on the structure strategy, the modeling of this pattern was done by human tutors (Meyer & Poon, 2001). In ITSS, the IT is an animated pedagogical agent who models the strategy use by speaking to the students in an animated Flash movie. As the IT speaks, the screen highlights segments to which he is referring and then prompts the learners to try it out themselves.

To motivate learners, we used a real human voice behind the animated pedagogical agent. Also, the reading passages were chosen to reflect what our middle school students were interested in reading. For example, on the basis of the input of two focus groups with middle school students, new passages were added to those used by Meyer et al. (2002) about famous athletes such as Hank Aaron, Babe Ruth, Andre Agassi, Marion Jones, Michelle Kwan, and others. We used multimedia design principles (Mayer, 2001) that promoted temporal contiguity, coherence, modality, and redundancy. The students listened to the narration by the IT while reading the highlighted terms on the page. The animations were focused toward highlighting critical parts of the passages. The screen design was lean, with every inch dedicated to aspects of affording a "reading" environment. The book shown in Figure 14.2 shows the reading-orienting interface.

Interaction Cycle Step 2: Practice

In an ideal situation, all learners will actively participate in understanding how to apply the structure strategy. Research on multimedia learning shows that students who manipulate simulations, interact with systems, and control their practice are able to apply their learning in a variety of transfer situations (Mayer & Moreno, 1998). On the other hand, Jacobson, Maouri, Mishra, and Kolar (1995) and McKeague (1996) have shown that students lack the metacognitive skill in gauging their understanding and controlling their navigational behaviors to improve learning. They both found that students using linear formats of hypertext learned more than counterparts who were given the option of selecting parts of the text to read.

ITSS is designed to foster the use of the structure strategy in fifth- through seventh-grade students and accommodate two types of learners: (a) students who need the linear type of structured format, which forces them to read and listen to all information, and (b) students with high metacognitive abilities that allow them to wisely choose how much help they need. To achieve this goal, the students have to practice using the strategy, internalize the critical concepts and processes, and be able to apply the strategy in diverse settings. The diverse settings allow the students to abstract the approach and use it in multiple domains. Nine practice tasks were designed to help students learn and apply the structure strategy:

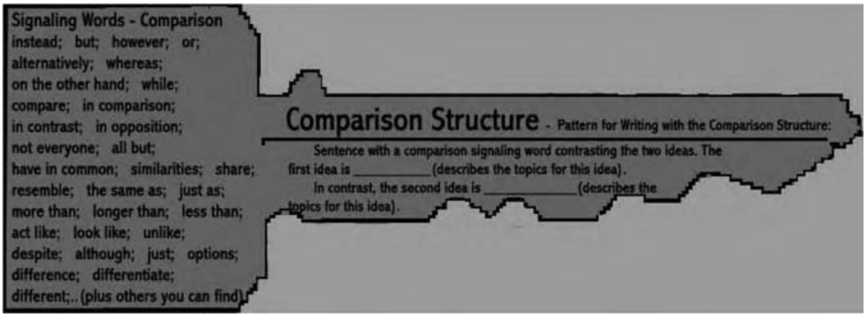


Figure 14.3 Comparison key back showing the signaling words.

1. Writing the name of the structure based on the signaling words in the passage. For example, if the student found the signaling word *different*, he or she wrote the structure: comparison.
2. Clicking on signaling words (with an available "key" to signaling words and expandable signaling word table within ITSS for each structure). The back of the key for the comparison structure is show in Figure 14.3. The key for each structure was a laminated learning aid provided for each of the five structures when the structure was initially introduced to the student and added to a key ring that was available for consultation during the lessons. The keys became a part of the program after Meyer et al. (2002) discovered that fifth-grade students had difficulty comparing information on two windows on a computer (i.e., a text in one window and signaling words in another).
3. Writing main ideas for the text passages they read. The front of the key showing the pattern for writing the main idea for a comparison passage is shown in Figure 14.4. In the first half of ITSS lessons for each structure, the student practiced writing the main idea while the passage was still on the screen. As students progressed, they had to write the main idea without having the passage available.
4. Writing a complete recall of the passage. The recall was written with the main idea the student had constructed earlier available to aid memory. The key (see Figure 14.3) shows the basic pattern to use in writing a recall with the comparison structure.
5. Filling in a tree diagram showing the ideas being compared, the problem and its solution or a cause and its effect, or the sequence of steps in a process or timeline.

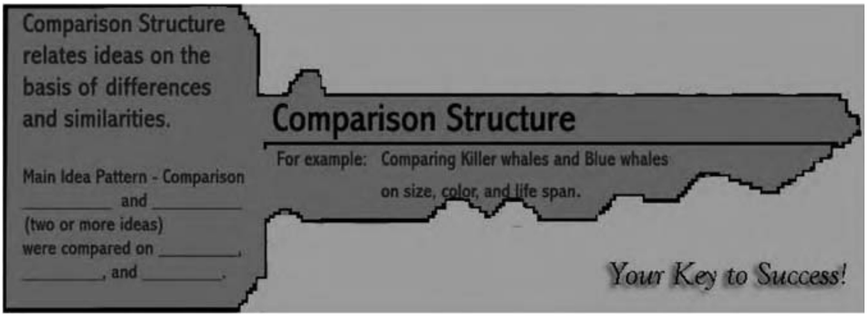


Figure 14.4 Comparison key front showing the main idea pattern.

6. Clicking on answers for multiple-choice questions. Multiple-choice questions were limited to a few lessons because we focused on students' construction of main ideas and recall of text information.
7. Creating titles for passages by the students.
8. Creating their own passages given signaling words and some general themes.
9. Correcting the work provided by the IT of other students' performances in the writing of main ideas or recalls.

These practice approaches were again implemented applying the findings of relevant research.

Interaction Cycle Step 3: Assessment

In the context of intelligent tutoring systems, assessment of student responses is the most critical component to help create the student cognitive model, identify the feedback given to the student, and generate the next move of the system. ITSS uses a two-step assessment of student responses. First, the responses (especially the full recall answers) are checked against a parse tree based on the propositional analysis in the content structures. This method also included checking the spelling and a synonym processor for the responses.

Second, the system checks the student response against an LSI system populated with data from previous research studies conducted by Meyer et al. (2002) and modeling from the research team. Both assessment methods scored the recall for main idea, structure, signaling words, and details. The LSI system also contains scores on top-level structure and intrusions. The scores and other measures, such as the "gaming" factors, blank responses, paragraphs, and number of tries at the question, are combined to find the type of feedback to be given to the student.

Interaction Cycle Step 4: Feedback

Feedback to the student is the final and critical step in helping support the learner and motivating them. The research on human tutors has shown improvements of two standard deviations in learning; most improvements are attributed to the ability of the tutor to adapt, scaffold, and provide feedback to the learner (Bloom, 1984). Corbett and Anderson (1991) showed that the time to complete lessons was significantly reduced with immediate feedback compared with no feedback. Anderson et al. (1995) reported that students learned better and faster with effective feedback. Koedinger (2002) presented feedback tactics that can be used with different topics. For example, the competence of the student, whether the topic had been mentioned before, and whether there was a student error were contributing factors to how the tutor would respond to the student.

ITSS uses a more formal feedback approach based on the student model created by using the database and LSI. The system was designed to provide advanced or simple feedback depending on the type of question and student response. Examples of advanced feedback for the first attempt at answering the question was "Your signaling words and main idea are correct; please add more details" or "Please check your comparison key for signaling words and try clicking on them again." Advanced feedback for the second attempt was also similar. At the third attempt, however, the students were shown a pop-up window with the correct answer. In the simple feedback condition, students were given a "good job" or "try again" type of feedback. The IT narrated the feedback to them, and in some conditions the pop-up window showed additional hints.

ITSS INSTRUCTION WITH FIFTH- AND SEVENTH-GRADE STUDENTS

We recently tested whether fifth- and seventh-grade students could learn the structure strategy by means of ITSS (Meyer, Wijekumar, & Middlemiss, 2005; Meyer et al., 2006). Students used ITSS during assigned class periods of 40 to 45 minutes twice a week for 6 months. We examined different feedback and motivation conditions in delivering ITSS. We compared a version of ITSS in which the system selected texts for practice lessons versus one in which the student made a choice between two topics of texts for each practice lesson. (See

Tables 14.1–14.3 and compare the columns with passage topics for standard lessons and parallel texts for topic choice.) Also, we compared the tutor giving minimal feedback to students' performances of "good," "try again," or "thank you" versus substantial and specific feedback from the IT. Students were stratified on proficiency of reading comprehension and then randomly assigned to the conditions, including conditions for counterbalanced testing materials over the pretest and posttest. In the experimental design pairs of students (one with choice and one without choice) read the same practice passages. One member of the pair always selected the topic of the practice text from two possible topics (see Tables 14.1-14.3 for text choices) while the other member of the pair read the topics previously selected by his or her linked pair. Choices for the comparison practice texts were made just before the 12 lessons about the comparison structure. With a sample of nearly all fifth-grade students in one elementary school, choice had an effect on recall produced at Lesson 10, a formative evaluation lesson (effect size = 0.57). This effect was not maintained on other formative evaluations or on the posttest. Thus, there was some evidence that choice of text topics could influence motivation in a particular lesson, but the effect was of short duration.

More specific feedback boosted performance in writing main ideas (specifically, listing the issues compared) on a formative evaluation text and performance of fifth-grade students (effect size = 0.74) and seventh-grade students (effect size = 0.36) on a standardized test of reading comprehension. Substantial differences between pretest and posttest performances were found on the Gray Silent Reading Test (Wiederholt & Blalock, 2000) for participants who received specific feedback (effect size = 0.48). Specificity of feedback (advanced vs. simple) did not affect performance on experimenter-designed testing materials delivered during the posttest. Significant differences between pretest and posttest performances were found on a researcher-designed measure of structure strategy use (effect size = 1.23), inclusion of the most important information in main idea statements (effect size = 0.58), amount of information remembered (effect size = 0.73), cloze performance on signaling words (effect size = 0.67), answers to main idea questions (effect size = 0.43), and the Motivation to Read profile (effect size = 0.21; Gambrell, Palmer, Codling, & Mazzoni, 1996).

We also identified reading related metacognitive strategies used by fifth- and seventh-graders, motivational factors pertinent to middle school students, and suggestions from students on making ITSS adaptable to their preferences (i.e., adding games). We interviewed 15 students using maximum variation sampling. The students interviewed varied on their reading ability, how many lessons they had completed, performance on each lesson, average number of attempts for each question, and performance on clipboard tests. The results showed that all interviewed students were unaware of the five text structures before ITSS training, although some high-ability learners were aware of the problem and solution structure but had never been formally introduced to using it as a method to organize their reading and writing. One high-ability student commented:

> All the language arts teachers and texts ask you to underline the important concepts or ask questions of yourself about the passage, but all we have experiences on are reading stories, this method is actually focused towards the type of readings we do in all other textbooks.

Another, extremely high-ability seventh-grade student explained, "I know about cause and effect and problem and solution structures just from reading and abstracting ideas. Now [via ITSS] I feel there is a complete set of cards to the deck I had started."

Results from an analysis of the ITSS logs revealed that some students were gaming the system by typing in blank responses to see how many tries they get. Eight students entered nonsensical responses such as a string of 5s for some of the questions. Some seventh-grade students who tried gaming the system initially in the comparison passages became more diligent on the next set of problem and solution lessons when they became more involved with some problems and author-favored solutions versus ill-favored solutions and even showed pretest-to-posttest gains in performance.

The structure strategy is a well-researched comprehension technique that is an essential foundation for any of the three threads of reading research. The new generation of intelligent learning technologies and research has afforded the creation of an interactive, Web-based, easily accessible, intelligent tutoring system, ITSS. ITSS has demonstrated success in improving reading comprehension for middle school students. The findings pointed to ITSS with elaborated feedback being particularly helpful for boosting standardized reading test scores of fifth- and seventh-grade readers. These practical and substantial improvements also were robust for struggling seventh-grade readers whose initial standardized reading test scores fell at least two grades below grade level.

ACKNOWLEDGMENTS

This research was supported in part by a Reading Comprehension grant from the U.S. Department of Education Institute of Education Sciences. More information on our project is available at http://itss.br.psu.edu/

REFERENCES

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, *4*, 167–207.

Aristotle. (1960) (L. Cooper, Trans.). *The rhetoric of Aristotle*. New York: Appleton-Century-Crofts.

Armbruster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly, 22,* 331–346.

Bartlett, B. J. (1978). *Top-level structure as an organizational strategy for recall of classroom text*. Unpublished doctoral dissertation, Arizona State University.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13,* 4–16.

Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, *19*, 727–752.

Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology*, *80*, 448–456.

Corbett, A. T., & Anderson, J. R. (1991, March). *Feedback control and learning to program with the CMU/LISP tutor*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Crothers, E. J. (1972). Memory structure and the recall of discourse. In J. B. Carroll & R. O. Freedle (Eds.), *Language comprehension and the acquisition of knowledge* (pp. 247–283). Washington, DC: Winston.

Dee-Lucas, D., & DiVesta, F. J. (1980). Learner-generated organizational aids: Effects on learning from text. *Journal of Educational Psychology, 72,* 304–311.

Duke, N. K., & Pearson, P. D. (2002). Effective practices for developing reading comprehension. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed., pp. 205–242). Newark, DE: International Reading Association.

Englert, C. S., & Hiebert, E. H. (1984). Children's developing awareness of text structures in expository materials. *Journal of Educational Psychology*, *76*, 65–74.

Fillmore, C. J. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory* (pp. 1–81). New York: Holt, Rinehart and Winston.

Gambrell, L. B., Palmer, B. M., Codling, R. M., & Massoni, S. A. (1996). Assessing motivation to read. *The Reading Teacher, 49,* 518–533.

Gordon, C. J. (1990). Contexts for expository text structure use. *Reading Research and Instruction*, *29*, 55–72.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.

Grimes, J. E. (1975). *The thread of discourse*. The Hague, The Netherlands: Mouton.

Jacobson, M. J., Maouri, C., Mishra, P., & Kolar, C. (1995). Learning with hypertext learning environments: Theory, design, and research. *Journal of Educational Multimedia and Hypermedia*, *4*, 321–364.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 165–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. Ruddell & N. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1270–1328). Newark, DE: International Reading Association.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394.

Koedinger, K. R. (2001). Cognitive tutors as modeling tool and instructional model. In K. D. Forbus & P. J. Feltovich (Eds.), *Smart machines in education: The coming revolution in educational technology* (pp. 145–168). Menlo Park, CA: AAAI/MIT Press.

Lajoie, S. (Ed.). (2000). *Computers as cognitive tools: No more walls. Theory change, paradigm shifts, and their influence on the use of computers for instructional purposes* (Vol. II). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lexile Framework for Reading. (2005). Lexile analyzer. Retrieved March 13, 2005, from http://www.lexile.com/DesktopDefault.aspx?view=re&tabindex=2&tabid=31&tabpageid=358

Loman, N. L., & Mayer, R. E. (1983). Signaling techniques that increase the understandability of expository prose. *Journal of Educational Psychology*, *75*, 402–412.

Long, D. L., & Lea, R. B. (2005). Have we been searching for meaning in all the wrong places? Defining the "search after meaning" principle in comprehension. *Discourse Processes*, *39*, 279–298.

Lorch, R. F., Jr., & Lorch, E. P. (1985) Topic structure representation and text recall. *Journal of Educational Psychology, 77,* 137–148.

Lorch, R. F., Jr., & Lorch, E. P. (1996) Effects of organizational signals on free recall of expository text. *Journal of Educational Psychology, 88,* 36–48.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111–151.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge, England: Cambridge University Press.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*, 312–320.

McKeague, C. A. (1996). *Effects of hypertext structure and reading activity on novices' learning from text*. Unpublished doctoral dissertation, Pennsylvania State University.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440–466.

McNamara, D. S. (2004, January). *Coh-Metrix: Automated cohesion and coherences scores to predict test readability and facilitate comprehension*. Paper presented at the Reading Comprehension Principal Investigators Meeting of the Institute of Educational Sciences in Washington, DC.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, and Computers*, *36*, 222–233.

Meyer, B. J. F. (1971). *Idea units recalled from prose in relation to their position in the logical structure, importance, stability and order in the passage*. Unpublished master's thesis, Cornell University, Ithaca, NY.

Meyer, B. J. F. (1975). Identification of the structure of prose and its implications for the study of reading and memory. *Journal of Reading Behavior*, *7*, 7–47.

Meyer, B. J. F. (1999). The importance of text structure in everyday reading. In A. Ram & K. Moorman (Eds.), *Understanding language understanding: Computational models of reading* (pp. 227–252). Cambridge, MA: MIT Press.

Meyer, B. J. F. (2003). Text coherence and readability. *Topics in Language Disorders*, *23*, 204–224.

Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly*, *16*, 72–103.

Meyer, B. J. F., & Freedle, R. O. (1984). Effects of discourse type on recall. *American Educational Research Journal*, *21*, 121–143.

Meyer, B. J. F., & McConkie, G. W. (1973). What is recalled after hearing a passage? J*ournal of Educational Psychology*, *65*, 109–117.

Meyer, B. J. F., Middlemiss, W., Theodorou, E., Brezinski, K. L., McDougall, J., & Bartlett, B. J. (2002). Older adults tutoring fifth-grade children in the structure strategy via the Internet. *Journal of Educational Psychology*, *94*, 486–519.

Meyer, B. J. F., & Poon, L. W. (1997). Age differences in efficiency of reading comprehension from printed versus computer-displayed text. *Educational Gerontology*, *23*, 789–807.

Meyer, B. J. F., & Poon, L. W. (2001). Effects of structure strategy training and signaling on recall of text. *Journal of Educational Psychology*, *93*, 141–159.

Meyer, B. J. F., & Rice, G. E. (1982). The interaction of reader strategies and the organization of text. *Text, Interdisciplinary Journal for the Study of Discourse*, *2*, 155–192.

Meyer, B. J. F., & Rice, G. E. (1989). Prose processing in adulthood: The text, the reader, and the task. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and later life* (pp. 157–194). New York: Cambridge University Press.

Meyer, B. J. F., Talbot, A. P., Poon, L. W., & Johnson, M. M. (2001). Effects of structure strategy instruction on text recall of older African American Adults. In J. L. Harris, A. Kamhi, & K. E. Pollock (Eds.), *Literacy in African American communities* (pp. 233–263). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Meyer, B. J. F, Wijekumar, K., & Middlemiss, W. (2005, July). *Web-based intelligent tutoring for the structure strategy*. Report to the U.S. Department of Education Institute of Education Sciences.

Meyer, B. J. F., Wijekumar, K., Middlemiss, W., Meier, C., Spielvogel, J., & Higley, K. (2006, April). Impact of computer-delivered strategy instruction on student outcomes. In A. Sweet (Chair), *The many faces of strategy instruction*. Symposium conducted at the meeting of the American Educational Research Association, San Francisco.

Meyer, B. J. F., Young, C. J., & Bartlett, B. J. (1989). *Memory improved: Reading and memory enhancement across the life span through strategic text structures*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Myers, J. L., & O'Brien, E. J. (1989). Accessing the discourse representation during reading. *Discourse Processes*, *26*, 131–158.

Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology*, *76*, 1239–1252.

Polley, R. R. (1994). *Facilitating recall through awareness of text structure*. Unpublished master's thesis, University of Kentucky.

Pressley, M., & McCormick, C. B. (1995). *Advanced educational psychology for educators, researchers, and policymakers*. New York: HarperCollins.

Raphael, T. E., & Kirschner, B. M. (1985). *The effects of instruction in comparison/ to compare/text structure on sixth-grade students' reading comprehension and writing products* (Research Series No. 161). East Lansing: Michigan State University, Institute for Research on Teaching.

Richgels, D. J., McGee, L. M., Lomax, R. G., & Sheard, C. (1987). Awareness of four text structures: Effects in recall of expository text. *Reading Research Quarterly*, *22*, 177–196.

Samuels, S. J., Tennyson, R., Sax, L., Mulcahy, P., Schermer, N., & Hajovy, H. (1988). Adults' use of text structure in the recall of a scientific journal article. *Journal of Educational Research*, *81*, 171–174.

Sanders, T. J. M., & Noordman, L. G. M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, *29*, 37–60.

Slater, W. H., Graves, M. F., & Piche, G. L. (1985). Effects of structural organizers on ninth-grade students' comprehension and recall of four patterns of expository text. *Reading Research Quarterly*, *20*, 189–202.

Taylor, B. M., & Beach, R. W (1984). The effects of text structure instruction on middle grade students' comprehension and production of expository text. *Reading Research Quarterly*, *19*, 134–146.

Theodorou, E. (2006). *Comparing the effects of learning the structure strategy via Web-based training or classroom training on the recall of near and far transfer texts*. Unpublished doctoral dissertation, Pennsylvania State University.

Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition*, *16*, 283–298.

van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, *39*, 299–316.

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. In S. R. Goldman & H. van Oostendorp (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wiederholt, J., & Blalock, G., (2000). *Gray silent reading tests*. Austin, TX: Pro-ed.

Weisberg, R., & Balajthy, E. (1989). Transfer effects of instructing poor readers to recognize expository text structure. In S. McCormick & J. Zutell (Eds.), *Cognitive and social perspectives for literacy research and instruction: Thirty-eighth yearbook of the National Reading Conference* (pp. 279–285). Chicago: National Reading Conference.

Williams, J. P., Hall, K. M., & Lauer, K. D. (2004). Teaching expository text structure to young at-risk learners: Building the basics of comprehension instruction. *Exceptionality*, *12*, 129–144.

Williams, J. P., Hall, K. M., Lauer, K. D., Stafford, K. B., DeSisto, L. A., & deCani, J. S. (2005). Expository text comprehension in the primary grade classroom. *Journal of Educational Psychology*, *97*, 538–550.

15

Guided Practice in Technology-Based Summary Writing

Donna Caccamise, Marita Franzke, Angela Eckhoff, Eileen Kintsch, and Walter Kintsch
University of Colorado, Boulder

The computer tutor Summary Street addresses the problem of passive comprehension strategies by providing content feedback on students' written summaries. The design of this tool is motivated by the belief that articulating their understanding in their own words helps students lay the conceptual foundation to learn about new subject areas. Graphically presented feedback, based on latent semantic analysis, guides this process to ensure adequate content coverage within given length constraints and the avoidance of redundant, irrelevant, overly detailed, or plagiarized information. In this chapter, we review our experimental findings, showing not only improved summary writing when using the tool but also transfer to independent summary writing, among middle school students. Students who practiced summarizing without guidance did not perform as well. Moreover, Summary Street users also scored higher on comprehension test items that required gist-level understanding. Results of a verbal report study with college students indicated better metacognitive awareness of appropriate strategies among those who composed summaries with Summary Street. Thus, our suggestion that better macroprocessing strategies may emerge from guided practice receives support from these findings.

Current psychological theories view text comprehension not as a single, unitary process but as consisting of several levels of processing (Kintsch, 1998): A coherent representation of the text meaning must be constructed from the printed words and sentences, at both the local and global level. Because no (readable) text is fully explicit, inferences are required to fill in gaps between individual propositions, the meaning units of language, and between groups of propositions. Further inferences are required to arrive at an interpretation of the text in the light of one's personal knowledge and experience, the level of meaning representation termed the *situation model* by Kintsch (1998). Hence, there is more to reading expertise than successfully decoding the printed words and sentences and many skills to master on the pathway to reading expertise.

A key component of successful comprehension is the degree to which readers actively work at constructing the meaning of what they are reading. Although comprehension of easy texts, such as stories and texts about familiar content, proceeds fairly automatically, effortful, strategic processing is crucial to learning from the more complex instructional texts encountered in later school years. Active meaning construction involves being tuned into one's own thinking, always alert to whether comprehension is proceeding smoothly. In the latter case, the reader must infer missing information, such as pronoun or synonym referents and various cohesive links to complete the coherence structure of the text. Deeper level inferences, such as thinking of analogies, examples, and questions, or re-explaining a sentence or passage in one's own words, serve to link the incoming information with what is already known. This kind of activity personalizes the meaning and results in really understanding the situation depicted in a text. It is this kind of deep-level processing that creates knowledge that can be rapidly and flexibly accessed as the situation requires (Kintsch & Kintsch, 2005).

In keeping with the multileveled view of reading comprehension, our instructional interventions now actively address the problem of readers who, despite good decoding skills, nevertheless fail to comprehend deeply, that is, who fail to construct an accurate and coherent textbase and well-elaborated situation model (Kintsch, 1998; van Dijk & Kintsch, 1983). Interventions that provide instruction on comprehension strategies, targeting either individual readers or whole classes of learners, are the focus of most of the contributions to the present volume. A somewhat different approach is pursued in our work, for not only is it important to know about good comprehension strategies, but learners must also have many opportunities to practice using them. Like other expert skills, reading expertise develops across many years of intensive, deliberate practice, guided by feedback (Ericsson & Kintsch, 1995). Indeed, an interesting theoretical question is the extent to which knowledge about reading strategies may emerge from practice in richly varied and supportive contexts, much as young children do when they acquire language from their surroundings. Most people are able to induce the linguistic regularities in their native language and express themselves correctly with little awareness of the underlying rules described by linguists. Likewise, expert readers may be largely unaware of their reading processes, except when the text is exceptionally difficult and the subject matter is unfamiliar. Although trying to grasp the author's meaning in such situations may require conscious, effortful processing, for the most part expert readers probably do not realize that they are paraphrasing or self-explaining, finding analogies or predicting outcomes, and so on. They just do it. Expert readers, however, do realize when there is a problem, which is often not the case with less skilled readers (Brown, Bransford, Ferrara, & Campione, 1983). The faulty metacognitive skills of nonexpert readers are moreover compounded by a tendency to read in a passive, uninvolved, or unengaged manner. Although little effort is required for easy-to-comprehend text, such as a straightforward narrative, passive reading of complex, informational texts will not support deep understanding of the content.

The problem of passive comprehension strategies thus becomes a major impasse to academic progress in the middle and high school years, as Bereiter and Scardamalia (1985) noted in their discussion of the problem of "inert knowledge." Changing barely proficient readers into expert readers is an important goal when we consider that most U.S. teens are not expert readers. The national report card (National Center for Education Statistics, 2004) indicated about half of 8th- and 12th-grade students score at or below a basic level of reading comprehension. Although some of this outcome can be attributed to lack of decoding skills, it has been estimated that as many as 60% of readers who fall into the bottom half of these measures lack proper comprehension skills that enable in-depth content analysis and processing. This profile of our nation's youth is of particular concern because we expect them to achieve higher levels of academic accomplishments that require understanding and learning complex issues in areas of science and technology as these students keep pace with the demands of the 21st century (Caccamise & Snyder, 2005). This problem is the central focus of the strategy interventions and technology support described in the present volume.

Summary Street is computer software that supports the construction of written summaries. The design of this tool is guided by the belief that summary writing—articulating understanding in one's own words—can help students lay the conceptual foundation needed to learn about new subject areas. Summary Street can also alert teachers to gaps in their students' understanding, which can be addressed in instruction. Moreover, repeated practice in summary writing with supportive feedback should, over time, affect students' reading strategies in a positive manner, essentially leading them to become more active, engaged readers. This is the overarching goal of the software we are describing here.

In this chapter, we present some research results in support of guided practice as a route to reading expertise. The notion of guided practice is hardly new; indeed, it has long been engrained in parenting, mentoring, and teaching practices across human societies. A lot of what happens in one-on-one tutoring session is guided practice, with effects on learning that are much greater than what can be accomplished in a classroom situation (e.g., Bloom, 1984; Cohen, Kulik, & Kulik, 1982; Graesser, Person, & Magliano, 1995). Thus, recent analyses of human tutoring lend support to our broadly held notions about the value of guided practice (e.g., Aleven & Koedinger, 2002; Azevedo, 2005; Cromley & Azevedo, 2005; Graesser, Bowers, Hacker, & Person, 1997; Graesser et al., 1995; Hume, Michael, Rovick, & Evens, 1996; Merrill, Reiser, Ranney, & Trafton, 1992).

The effectiveness of guided practice unfortunately is severely limited by the constraints of typical school classrooms. With too many students, too much material to cover, and too many other demands competing for the teachers' time, most teachers rely on supplementary activities, often in the form of workbook exercises, to provide additional, structured practice on taught concepts. Of course, all forms of guided practice work best if they are tailored to a student's particular need, that is, geared to the student's present level of understanding. Workbook exercises are notorious for decomposing learning material into small, discrete bits that fit the easily scored, objective-question format. Feedback is limited to "right" or "wrong" and is generally delivered well after the exercise has been completed. Consequently, students rarely learn why a given answer is wrong and rarely get a second attempt to improve on it. These issues have been with us for a very long time, but today's technological breakthroughs, along with a better, more differentiated view of comprehension and learning processes (e.g., Kintsch, 1998), are beginning to inform the design of sophisticated computer-based tools and innovative instructional practices that support true learning.

ROLE OF FEEDBACK IN SUMMARY STREET

Researchers and educators alike have long been convinced of the value of providing feedback on students' work as soon as possible after the activity is completed (e.g., Druckman & Bjork, 1994). However, it is a difficult-to-achieve goal in real world classrooms, although teachers do a remarkable job of providing rapid feedback for activities with easy-to-score, closed-entry answers, or if the feedback can be delivered to the classroom as a whole. The former kind of feedback typically lets the student know whether his or her answer was right or wrong, whereas the latter kind of feedback tends to be at a fairly general level, addressing problems the group as a whole may be experiencing. Providing rapid, individually tailored feedback at a level appropriate to the problem at hand is probably an important reason why skilled tutors are so effective (e.g., Aleven & Koedinger, 2002; Cromley & Azevedo, 2005; Kulik & Kulik, 1988; Merrill et al., 1992). Given the large classes and limited resources in most schools, it is rare that teachers are able to provide individualized help on an open-ended task; hence, students receive little guidance through successive drafts of their writing. The computer tutor Summary Street, which is the focus of this chapter, does a remarkable job of supplementing the teachers' instruction on an open-ended writing task, namely, summarizing. A summary submitted for feedback in Summary Street goes via Internet connection to the Pearson Knowledge-Technologies Web site (http://www.pearsonkt.com), where the content of the summary is analyzed by *latent semantic analysis* (LSA), a machine learning method that models human semantic knowledge. It takes as its input a large corpus of text, representative of the texts students would have encountered during the course of 14 years of schooling, and extracts from that a representation of the semantic content of a submitted text or document. It can judge the similarity in meaning of two texts as well as human judges can, regardless of the actual words used. The rating of similarity is expressed as a cosine, similar to a correlation coefficient. It is important to note that similarity ratings by LSA do not depend on the presence or absence of particular key words from the source text; instead, the method provides an estimate of the overall meaning of the content that is conveyed. (More details about the method are available in Landauer, 1998, and Landauer & Dumais, 1997). This capability to compare the meaning similarity of texts is the basis for the feedback that Summary Street provides on several aspects of the content of a student's written summary.

The feedback, which is returned almost instantaneously, is in the form of a graphic display, shown in the example in Figure 15.1, which indicates whether the main topics of the source text have been adequately covered by the summary. Each horizontal bar corresponds to one of the topic sections of the source text, labeled with the headings used in the source text. The student's goal is to get the bar past a vertical line representing the threshold (an LSA cosine) for content coverage for each section. The bars, initially orange, turn green when the threshold has been reached. A vertical bar on the right shows whether the summary length is within the appropriate range, as determined by the teacher. This bar likewise turns from orange to green when the summary is of the right length. A set of tools located at the bottom left corner provides links to additional functions that check for possible problem sentences: sentences

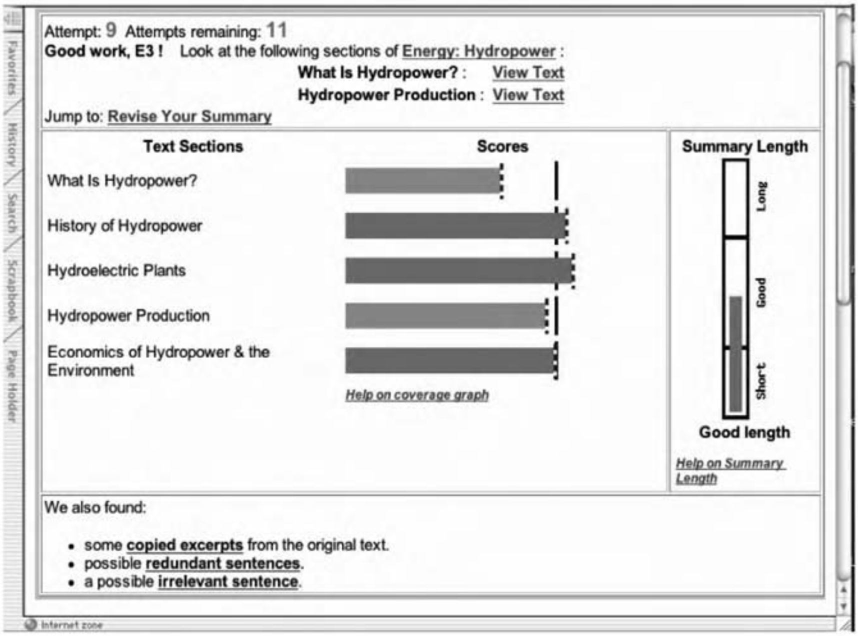


Figure 15.1 Summary Street window, showing feedback screen for "Hydropower": Horizontal bars, labeled with section headings from the text, denote content coverage. The bars turn from orange to green when the threshold, indicated by the black vertical line, is reached. The vertical bar on the right is a length indicator, which is green if summary is in the appropriate range and orange if it is too short or too long. Additional functions are shown at the bottom left side of the screen.

that are irrelevant, that are redundant with other sentences in the text, ones that are plagiarized from the source text, as well as a spelling help (not shown in the figure). Interested readers may visit the Colorado Literacy Tutor Web site (http://www.colit.org) to try out Summary Street as a guest user.[1] A detailed description of the tutor is available in Wade-Stein and Kintsch's (2004) article.

[1]Summary Street® is now available as a commercial tool by Pearson Knowledge-Technologies (http://www.pearsonkt.com).

We would like to stress here that the feedback provided by Summary Street is rich in that it deals with the meaning conveyed by students' writing, but except for spelling it does not address issues with writing mechanics (e.g., grammar, sentence structure, or punctuation). The feedback is multifaceted in that it addresses a number of content-based writing problems, problems with repeated information, information that is irrelevant or unimportant, and wording that is too close to the source text. Yet the guidance is relatively nondirective. It merely suggests potential problems with the writing, leaving it up to students to fix them as best they can or to ignore a flagged problem altogether. Indeed, sometimes sentences flagged as redundant are needed to maintain coherence, or they may describe concepts requiring multiword expressions, which were flagged as plagiarized, or they may be very brief sentences that have too little content to satisfy the LSA relevance criterion. Students are, of course, alerted to these issues and instructed that they must judge for themselves whether there is a problem and, if so, how to address it, keeping in mind whether their summary is of the appropriate length.

In our observations of classroom use of Summary Street, we find that students appreciate having a sense of control over the computer, using the feedback to make their own decisions, and treating the sometimes less-than-optimal flags as a challenge. Hence, our observations, together with the results of empirical trials with the system, are in line with recent analyses of human tutoring sessions on problem-solving tasks (e.g., Aleven & Koedinger, 2002; Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Graesser et al., 1995; Hume et al., 1996). These studies suggest that feedback that is neither too specific, in terms of supplying the right answer, nor so general (e.g., "poor," "fair," "good," "excellent") that the user doesn't know what to do next appears to be more useful to students because it indicates a possible path to the correct answer. This intermediate level of feedback tends to elicit more meaning-construction activity on the part of the student, whereas feedback that consists of an explanation or that simply provides the answer is more likely to terminate the student's effort. In the case of Summary Street feedback, we believe that simply pointing to possible problems appears to act indirectly as a scaffold for greater effort on the students' part to persist with the writing task.

The research described in the following sections of this chapter supports our claim that making students aware of problems with the content of their writing is highly effective with respect to improving the quality of the product. In addition, we present preliminary evidence that extensive practice in summary writing with Summary Street may transfer to reading comprehension and learning in general.

RESEARCH FINDINGS

In this section, we summarize research findings from three recent studies involving Summary Street. Although these studies span the range from large-scale evaluation to small laboratory studies, they all looked at Summary Street use over an extended period of use with students summarizing four or more texts. Earlier research has focused on briefer exposures to the technology (Wade-Stein & Kintsch, 2004) and found significant improvements in the summary quality of students using Summary Street compared with students who summarized texts without receiving feedback. More recently, our focus has shifted to seeking evidence of lasting benefits from the type of guided practice that Summary Street can provide. Specifically, these studies involve practice over multiple editing sessions summarizing four or more texts with the following three types of guidance: (a) feedback on whether the summaries adequately cover the information in the source text (main content feedback); (b) whether the content is expressed in the student's own words (through the plagiarism check); and (c) whether the writing avoids irrelevant, overly detailed, and redundant sentences (through the combination of checks on length, redundancy, and irrelevance). Study 1 provides evidence from several middle school classes of transfer of improved summarization skill to independent summary writing outside the context of Summary Street. In addition to replicating the positive effect on summary quality, Study 2 considers whether summary writing practice with Summary Street feedback improves students' sensitivity to important ideas in a text, as indicated by their superior performance on a reading comprehension post test. This result would support our claim that the tool encourages a more active stance toward reading that should benefit comprehension overall. Finally, Study 3 explores whether there is evidence of improved metacognitive awareness and strategy use. In the following sections, we address these issues in turn.

Study 1: Transfer to Independent Summary Writing

In our rollout of Summary Street into classrooms of the Colorado public school system, we are collecting evaluation data as part of an ongoing Interagency Educational Research Initiative grant administered by the National Science Foundation. The goal of this grant is to look at the elements that affect scalability and sustainability of scientifically based instructional tools for literacy. For the overall evaluation, normative student achievement data (statewide standards-based testing), nationally standardized comprehension test scores, specific quizzes that assess students' learning from text, and independently written summaries are collected at the beginning and end of the school year. We are using a matched control group approach in which experimental classes (at each class level) are matched with comparable control classes that do not use

Summary Street to assist in their reading and writing activities. The classrooms using Summary Street vary from elementary fifth-grade classes to high school ninth-grade classes, but demand has been highest in Grades 6 through 8, where students depend more on expository materials in their content area classes and where summary writing is practiced during language arts instruction. At the time of this writing, we have analyzed summarization data from six seventh- to ninth-grade experimental classrooms and five control classes that used Summary Street during spring 2004. On average, the students in the experimental classes had used Summary Street to summarize 4.9 texts. Control classes came from across the curriculum, in which texts served different purposes, such as a source of background information about science topics or incorporated into the reading and writing activities of language arts classes. In some of these control classes, students were also writing summaries but were not assisted by any feedback-driven tools.

As part of the overall evaluation of Summary Street, all the students in the experimental and control classes wrote a summary at the beginning and end of the school term. In the experimental classes, the Summary Street intervention was embedded in the regular curriculum. For these pre and post exercises, students read one of two texts ("Dust Bowl" or "Influenza"), which are similar in length (954 and 1,175 words, respectively) and reading level (12th grade). Both texts have four sections, all approximately equivalent in length. Reading and summary writing, using pencil and paper, took place during a single, 45-minute class period. Half of the students of each class received one text for the pretest, and the other half the other text in a randomized assignment. For the posttest, each student received the text he or she had not summarized during the pre-assessment.

Data from six experimental and five control classes were available for analysis at this time. Four of the experimental classes, and three of the control classes, were seventh-grade classes; one each in the experimental and control condition were eighth-grade classes; and one each in the experimental and control condition were ninth-grade classrooms. The classes were from a mixture of urban, suburban, and rural middle and high schools, but experimental and control classes were matched demographically. In total, there were 127 experimental and 116 control students in this sample; however, attrition rates reduced the sample to 80 students in the experimental group and 60 in the control group with usable data. To analyze the data, the transcribed summaries were submitted to Summary Street to obtain LSA cosines for content coverage of the source text.[2] These cosines were then compared to the threshold for each section of the source text to determine whether the

[2]Previous research has shown that scoring by LSA is as reliable as the scores of human graders (e.g., Landauer, Laham, & Foltz, 2003).

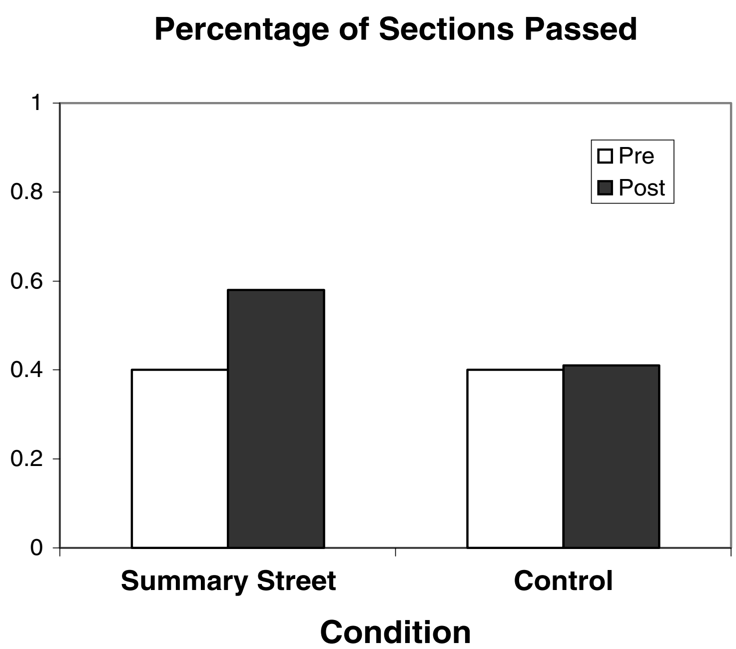


Figure 15.2 Mean percentage of text sections that passed threshold in pre- and post summaries as a function of group.

summary passed the content requirement for this section. A total passing score for each text was derived from the section scores (0, 1, 2, 3, or 4 sections passed corresponds to 0%, 25%, 50%, 75%, or 100%, respectively). As demonstrated in Figure 15.2, at the beginning of the spring 2004 term students passed only about 40% of the sections of the text when their presummaries were evaluated with Summary Street. At the end of the term, after practicing summary writing using Summary Street with an average of 4.9 texts, the experimental group's performance increased to passing 58% of the sections, whereas the control group remained unchanged, at 41%. This amounts to an effect size of $d = 0.38$, and the difference is statistically significant: $F(1, 139) = 4.9$, $p < .05$.

Another way of looking at the improvement in summary performance is in terms of the mean deviation from the section thresholds set by Summary Street. First, a deviation score for each of the four text sections was derived [(score – threshold)/threshold]. Then the average of the four deviation scores

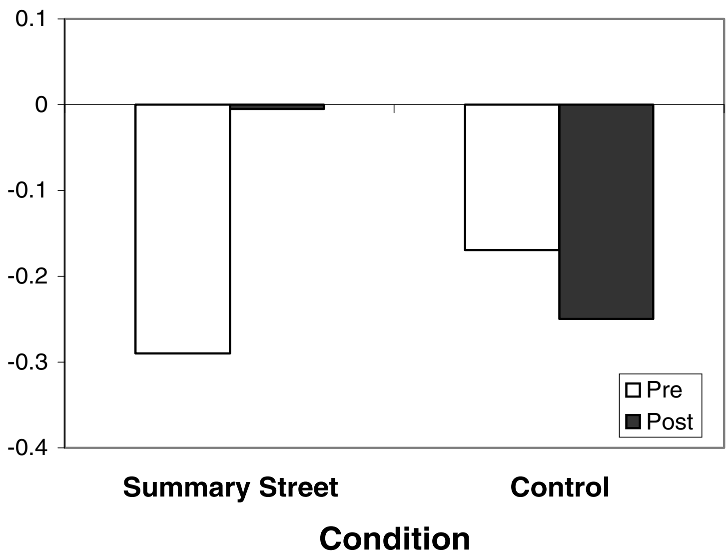


Figure 15.3 Deviation of mean scores from section thresholds in pre- and post summaries as a function of group.

per summary was used as a measure of how closely the content in a summary approximates the cosine thresholds for each text. A large negative deviation indicates that students did not include enough of the relevant semantic material in their summaries; small deviations indicate that the students included the right balance of important and detailed information. This finding is illustrated in Figure 15.3.

One can see that both the experimental and control groups started with quite sizable negative deviations from the thresholds (–0.17 and –0.29 mean cosines below the expected values, respectively). Although the control group performed slightly better on the pretest, this difference was not significant. On the posttest, however, the cosines of the experimental group almost matched the threshold after their practice with Summary Street (–0.005 mean cosines), indicating substantial improvement in the content coverage of their summaries. In contrast, the control group's performance on the post summary showed no improvement (–0.25 mean cosines). The effect size is $d = .67$, and the difference between conditions on the posttest is highly significant, $F(1, 139) = 15.4$, $p < .0001$.

It is clear that the guided practice that these seventh-, eighth-, and ninth-graders received during only one semester of Summary Street use, when summarizing approximately five texts, had a positive effect on the summaries written without this support at the end of the semester: Their summaries included more of the important content. As mentioned earlier, additional data are still being assembled from a large number of fifth- to ninth-grade classrooms. These data will include students' scores on national standardized tests and the state standards test, together with scores on researcher-designed quizzes administered after the pre- and postsummary assessments. These data sources will enable us to look for possible generalization to comprehension and learning from the improved ability to recognize and capture main points during summary writing.

Study 2: Guided Versus Unguided Practice

Although the large-scale evaluation design enables us to collect overall outcome data from control classes, which serves as a viable contrast to experimental classroom outcome data, it does not shed light on the effect of specific practices in the control classrooms. This gap in our knowledge is due to the fact that it is hard to control what writing activities students in these classes actually do; some teachers practice summary and essay writing, others only assign texts for classroom or homework reading.

To address this question—whether guided practice (frequent feedback on the content of summaries) is superior to unguided practice (summary writing without receiving feedback)—we conducted a smaller scale classroom study, using four eighth-grade language arts classes in a suburban Colorado middle school taught by the same teacher. For a detailed description of this study, see Franzke, Kintsch, Caccamise, Johnson, and Dooley (2005).

Working with one teacher and a relatively small number of students, who were randomly assigned to condition within each class, allowed us maximal control over the activities in the experimental and the control conditions. For 4 weeks, during two 45-minute sessions per week, students in both conditions wrote summaries of the same texts in the same order. Both experimental and control students were given the same instructions for summary writing at the beginning of this period. The instructions, taken from the guidelines to summary writing in the introductory page of Summary Street, were provided in written form to all students to refer to as desired. All students used the same computer laboratory for this exercise, where the experimental students logged into Summary Street while the control students simply wrote their summaries using Microsoft Word and saved them on a group server. The first and the last of the eight sessions were used for pre- and posttesting. During the summary writing sessions, all students read and summarized texts that were provided in a booklet. The booklet contained a mixture of short- to medium-length expository texts and stories, organized in order of difficulty. Students were instructed to work at their own pace and summarize as many texts as they could, making a good quality summary their primary objective. During the pre- and posttest, we administered a test battery of comprehension items from retired CSAP (Colorado State Assessment Program) materials, a standards-based test. Additionally, we had access to the students' CSAP and Scholastic Reading Inventory (Scholastic, 1999) test scores collected by the school district during the previous school year. The analyses are based on 115 eighth-grade students, approximately half in each condition.

Students in both the control and experimental conditions made similar progress in the number of texts they summarized during the six practice sessions. Experimental participants summarized an average of 6.55 (SD = 1.72) texts, and control participants summarized 5.75 (SD = 2.40); a t test comparing these population means was not significant, $t(44) = -1.55$, $p = .12$. Thus, students in both conditions had contributed comparable efforts to the summarization tasks.

Because of the time-consuming pre- and postcomprehension tests, students in this study were not asked to provide pre and post summaries. However, to gain some insight into the quality of the summaries produced during the practice, the summaries of four expository texts composed by all students were scored independently. Because they introduce a new confound, the two narrative texts were omitted from the analysis. The texts had been presented in order of difficulty, such that the first one had only one section and was 696 words long, and the last one had four sections with a total of 954 words. Two graders provided separate scores for overall quality, content coverage, organization, mechanics, exclusion of unnecessary detail, style, and plagiarism. As students progressed from easier to more difficult texts, experimental students outperformed the control students in all but two of these measures: (a) mechanics and (b) plagiarism. Our discussion here will focus on the content scores and the detail scores (whether the summaries avoided excessive detail); a more detailed description of the results is available in Franzke et al. (2005).

The content score measured the degree to which each section of a summary included the relevant information. These scores ranged from 2 (*good content coverage*) to 0 (*no relevant content included*). Figure 15.4 shows that with increasing difficulty and length of the original articles, the experimental participants were able first to increase and then maintain their performance across all four summaries, whereas the control participant quickly lost track of the important subject matter. An analysis of variance (ANOVA) of the content scores,

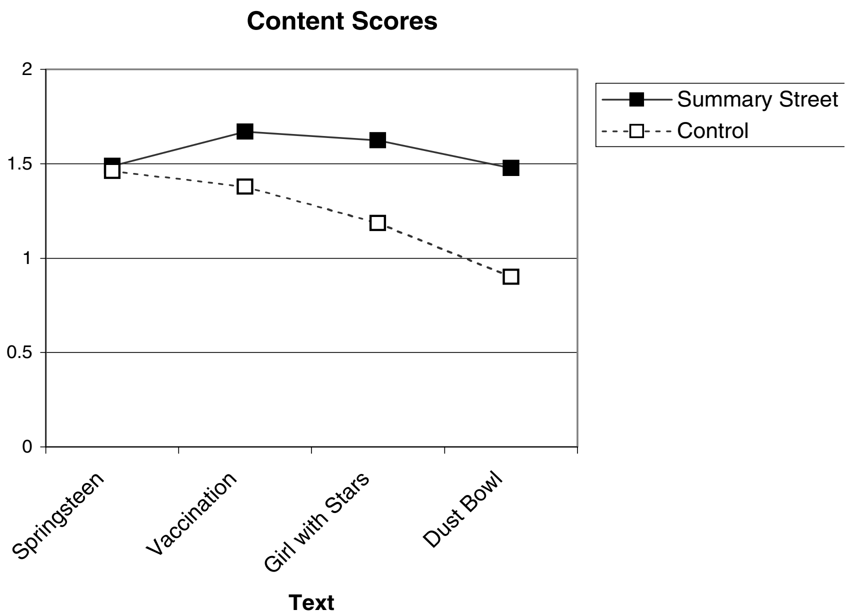


Figure 15.4 Mean content scores of summaries of four texts as a function of group (adapted from Franzke et al., 2005).

with condition and text as factors, revealed a highly significant main effect of condition. Expressed in effect sizes, for the easiest text, $d = 0.03$, but for the most difficult text, the effect size increases to $d = 0.82$.

One interpretation of the data from the evaluation study reported in the previous section could be that, with the help of Summary Street, students simply write more and therefore by chance also include more relevant information. However, the data from the present study suggest that students in the experimental condition were more selective in the content they included than were the control students. Summaries were scored from 0 to 5 points for exclusion of unnecessary detail. A high score (5) on this measure indicates that the summary had little detailed information, whereas a low score (1) indicates the inclusion of too many details. Figure 15.5 shows that students in the experimental condition increasingly learned to exclude such detail, whereas the control participants apparently did not learn this lesson when practicing summary writing without feedback. The ANOVA for these exclusion of detail scores revealed a highly significant interaction between text and condition, $F(2, 53) = 7.53$, $p < .01$. Again, the effect size for the easiest text is quite small

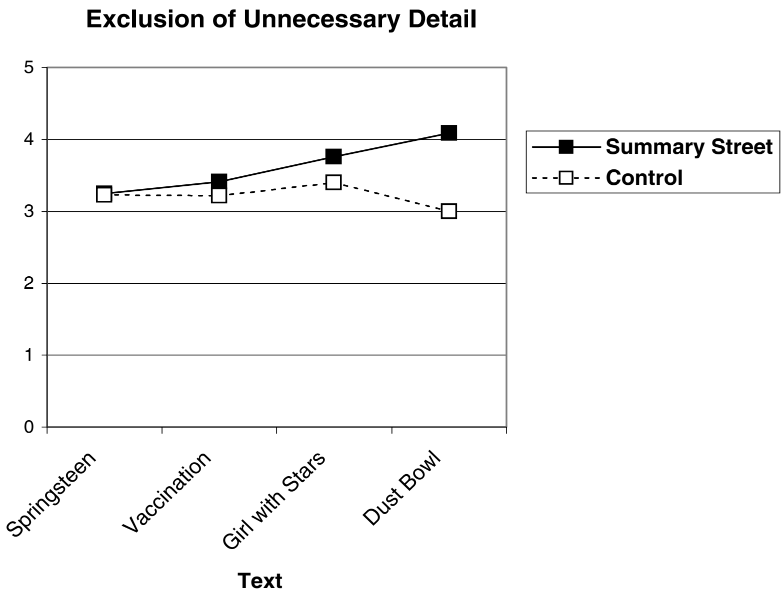


Figure 15.5 Mean scores for exclusion of unnecessary detail in summaries of four texts as a function of group (adapted from Franzke et al., 2005).

(*d* = .06) but increases to *d* = .93 for the most difficult text. Practicing summary writing when guided by Summary Street feedback apparently helped the experimental students not only to focus on relevant content but also to avoid less relevant, detailed content in their summaries.

When student ability was entered as another variable into the analysis of the data, it became apparent that especially students of lower and average ability were benefiting from the practice with Summary Street. An ANOVA showed a significant triple interaction among text, condition, and comprehension ability level (as determined from school administered CSAP tests) for content, as well as for a number of the other scores. Figure 15.6 shows the comparison of high-performing students with the average of the medium- and low-performing students. The difference between the two groups is significant according contrast tests, $F(3, 53) = 4.47$, $p < .05$. The effect size for high-performing students for the most difficult text is $d = 0.54$ and for the low- and medium-performing students $d = 1.5$. The analysis of content scores shown in Figure 15.6 suggests that guided practice in the form of immediate feedback that identified content problems with the writing was especially beneficial to students who most needed the help.

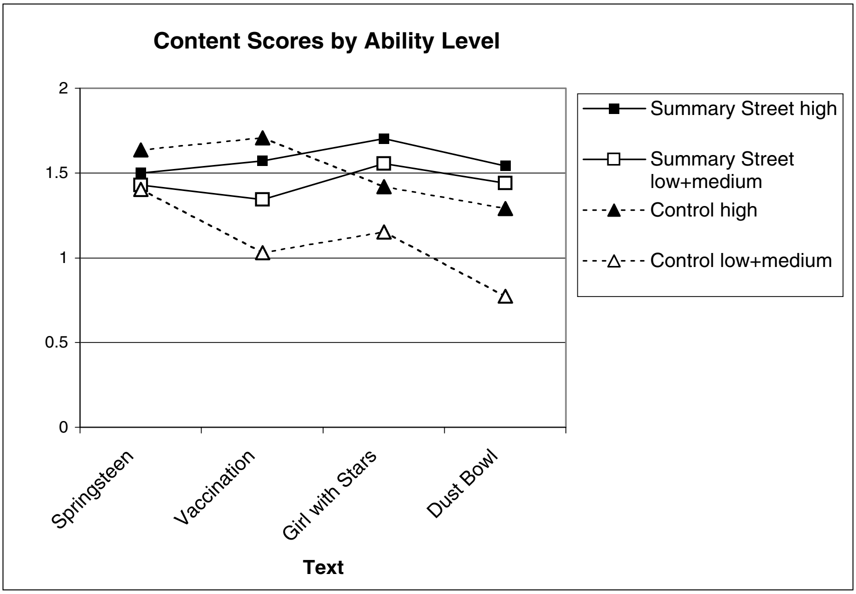


Figure 15.6 Mean content scores in summaries of four texts as a function of group and ability level (adapted from Franzke et al., 2005).

Our analysis of the comprehension pre- and posttests examined students' overall test performance as well as their performance on specific types of test items, namely, questions that tapped summary, inference, fact finding, and vocabulary skills. The only item group that showed a significant improvement on postintervention scores was the summary items. An analysis of covariance using pretest performance as a covariate revealed a significant difference on postintervention performance, $F(2, 83) = 4.05$, $p < .05$. This amounted to an effect size of $d = 0.42$. Thus, we have an indication that repeated guided practice writing summaries with the type of feedback provided by Summary Street can lead to improved performance on comprehension test items that focus on gist-level understanding.

At the conclusion of this study, we had also conducted interviews with a small sample of students (10 experimental and 8 control). One of the questions we asked was "What, in your opinion, is most important in producing a good summary?" The control and experimental students were from the same classes, instructed by the same teacher, and had received exactly the same general introduction to summary writing at the beginning of the intervention. Nevertheless, the students who practiced with Summary Street verbalized important summarization strategies more frequently than the control students: Eighty percent of the experimental group mentioned the strategy "cover the main points" compared with only 50% of the control group, "keep it short" was mentioned by 50% of the experimental students but only 15% of the control group, "cover each paragraph" was mentioned by 25% of the experimental students versus 15% of the control students, and "know the facts" was mentioned by 30% of the experimental students versus 15% of the control students. Although these interview data were collected incidentally and are qualitative in nature, this finding raises the question whether guided practice with Summary Street might lead to heightened metacognitive awareness of appropriate summarization strategies. If so, these insights into how to deal with content problems in summary writing should carry over to improving comprehension in other contexts. In the following section we describe a study (Eckhoff, 2005) intended to shed more light on this question. Although this study considered only adult performance, positive effects on strategic thinking in this population would encourage further investigation of the effect with middle to high school students, who are the targeted population of Summary Street users.

Study 3: Metacognitive Processes During Summary Street Use

Metacognitive strategies have been identified as an important component of learning and have been much studied since the early, foundational work of Brown (e.g., Brown et al., 1983) and Flavell (e.g., 1979). In general, *metacognition* refers to thinking about one's own thinking in order to monitor progress toward a particular goal and to assume active control over the strategies needed to accomplish it (e.g., Flavell, 1979; Gourgey, 2001; Hacker, 1998). Metacognitive awareness is the hallmark of expertise in most domains and has been especially studied in tasks involving reading comprehension, learning from text or other media, essay writing, and various kinds of problem-solving tasks. Because metacognitive processes involve active, conscious mental activity, akin to problem solving, evidence that Summary Street contributes to metacognitive thinking would support our argument that the tool is beneficial to comprehension and learning in general.

In Eckhoff's (2005) study, a verbal report method was used to identify the kinds of metacognitive processes occurring during summary writing and revision and to determine whether feedback from Summary Street enhances this kind of thinking among accomplished adult readers. Twenty college undergraduates were paid to participate in four hour-long summary composing sessions. Participants were tested individually in a quiet office, with half randomly assigned to the Summary Street group and the other half assigned to a control group. Four texts on sources of energy (biomass, propane, hydropower, and coal) were read and summarized by participants in each condition, one text per session. The texts formed a sequence in terms of increasing word length (from 1,155 words to 2,064 words) and number of sections (four to five sections). All participants were instructed to write summaries that were between 15% and 25% of the length of the source text.

After a brief warm-up task to practice thinking aloud, all participants were provided with a short review of guidelines for summarization. They were then instructed to verbalize their thoughts as they composed and revised their summaries. The sessions were tape-recorded for later transcription and analysis. Participants in the experimental group were told to submit their initial summary to Summary Street for evaluation (content feedback and any of the other tools available) as many times as needed until the summary passed all content and length requirements. Participants in this group requested feedback an average of four times. The control group participants wrote and revised their summaries on Microsoft Word, with access to the grammar and spell-check tools available there. They were instructed to revise each summary at least three times. Because the texts were not too difficult, all participants were able to finish the reading and summary writing task within the allotted time.

Each think-aloud comment in the transcribed protocols was categorized according to three main strategy types identified by Cross and Paris (1988) as follows:

1. *Planning strategies* comprised comments pertaining to goal-setting and selecting particular strategies to reach them (e.g., note-taking and underlining), allocating time, rehearsing what to say, selecting particular strategies for condensing text (i.e., the macro-rules described by van Dijk & Kintsch, 1983).
2. *Evaluation strategies* refers to evaluating one's goals and performance and determining the relevance of particular content, such as weighing main ideas.
3. *Regulation strategies* include monitoring and redirecting one's activities during reading, writing, and revising, for example, rereading the text, rechecking notes and other records, asking for help, and revising the written text at the sentence or paragraph level.

Time frames were added to each transcript from research notes taken during the experimental session. The reliability between two raters who assigned the think-aloud comments to categories was $r = .81$ across all texts.

The results indicate that the majority of metacognitive strategies occurred during revision for all texts and treatment groups. Thus, the results reported here are limited to strategies that were used during that time period.

The think-aloud protocols showed evidence of metacognitive strategy use among all participants, but the total number among Summary Street users was

TABLE 15.1
Means and Standard Deviations for Frequency of Metacognitive Strategies, for Which Significant Differences Occurred

	Summary Street		*Control*		
Strategy	*M*	*SD*	*M*	*SD*	*F*
Planning strategies					
using macro-rules	1.9	1.8	0.5	0.85	*
Evaluation strategies					
monitoring task performance	19.8	4.4	8.3	4.1	***
Regulation strategies					
checking notes & records	1.1	1.4	0.0	0.0	**
transforming text	8.3	3.7	3.1	2.7	***

$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

significantly higher than for the control group: mean = 84.1 versus 61.1, respectively. Additional ANOVAs revealed significant differences between the two groups in each of the three main categories: (a) planning, (b) evaluation, and (c) monitoring. As shown in Table 15.1, significantly higher scores were found among Summary Street users for using macro-rules, monitoring task performance, reviewing notes and records, and transforming written text. Heightened awareness of appropriate strategic processes thus appears to characterize summary writing with the aid of Summary Street for this population of older students. The extent to which these metacognitive benefits also hold for younger students is a subject to be considered in future research. The informal data collected by Franzke et al. (2005) suggest that this may be the case.

CONCLUSION

At the beginning of this chapter, we posed the question of whether knowledge about higher level comprehension strategies emerges from extensive guided practice in performing activities like summarizing text. Our results at this point, although suggestive, have not yielded a conclusive answer. A consistent finding, however, in all our research is that feedback from Summary Street—feedback that points out problems with conveying the content of a source text—helps students compose a better quality product according to various measures, compared with students who do not receive this feedback (Franzke et al., 2005; Wade-Stein & Kintsch, 2004). As the findings of Study 2 demonstrate, these improvements tend to be largest for low- and medium-ability students and for difficult texts. In other words, Summary Street helps most in situations where the student would otherwise struggle to capture the most important content and to distinguish it from lesser details.

The knowledge about what makes a good summary apparently stays with the students who have used the tool over multiple practice sessions and transfers to situations in which no feedback is given, as shown by the postintervention summaries. Moreover, as Franzke et al. (2005) demonstrated, merely practicing summary writing without guidance does not lead to comparable improvement.

Does knowledge about how to summarize emerge from guided practice in summary writing? Results from the think-aloud study with college students (Eckhoff, 2005) indicate that the Summary Street feedback does encourage metacognitive reflectivity among these older students. Informal interview data (Franzke et al., 2005) suggest that this may also be the case with younger students. Finally, students who practiced summary writing with the software not only produced better summaries but also performed better on the postintervention comprehension test items (those that required a brief summary response) than students who practiced summary writing without guidance. This result certainly suggests that Summary Street feedback may exert a positive effect on macroprocessing in general.

Several characteristics of the feedback embedded in Summary Street may be important in this regard. First, the feedback is individually tailored to each student's summary. Furthermore, it is delivered almost instantly on request. Also important is the fact that the feedback is pitched at the right level, neither too general as to be useless, nor too specific; flagged sentences are not "wrong" but rather suggestions about what may need fixing. The bar graphs on the content feedback page display gaps in topic coverage. Whether the benefits of Summary Street are primarily due to a particular type of feedback or whether it is a joint function of all of them cannot be determined on the basis of the findings reported here. Neither can we estimate the effect of the gamelike interface that presents users with a concrete goal: to cover the content adequately within the assigned length constraints. It appears, however, that Summary Street does make it possible for many students—especially readers of moderate to lower ability—to summarize more difficult texts than they could otherwise handle. In general, students are challenged to find their own solutions while writing and revising within this supportive context, which may lead them to be more actively engaged in meaning construction than they would without this guidance.

Summary writing is a valuable learning activity, because it helps readers build a coherent textbase understanding, which is the foundation for true learning. The ability to summarize a text is an accurate indication of how well a text has been understood. This is because writing summaries depends on the same underlying processes that readers use to distill a gist representation of the text meaning—the macro-strategies described by van Dijk and Kintsch (1983). Summary Street guides students through the process of constructing an accurate and sufficiently elaborated representation of the text macrostructure.

Our results suggest that these macro-strategies need not be explicitly taught but can be induced if readers have sufficient opportunities to practice doing activities like summarizing text. Like so many other skills, the pathway to reading expertise requires a great deal of practice. However, as Ericsson (Ericsson, Krampe, & Tesch-Römer, 1993) has shown in a wide variety of domains, sheer amount of practice does not necessarily make one an expert. Practicing must be carried out in an intelligent, deliberate manner, guided by a skillful mentor. Lacking the availability of a sensitive human tutor, even a semi-intelligent agent like Summary Street can help students learn higher level writing and comprehension strategies through summarizing.

ACKNOWLEDGMENTS

The work described in this chapter was supported by a grant from the Interagency Educational Research Initiative of the National Science Foundation (030.05.0431B).

REFERENCES

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, *26*, 147–179.

Azevedo, R. (2005). Computer environments as metacognitive tools for enhancing learning. *Educational Psychologist*, *40*, 193–197.

Bereiter, C., & Scardamalia, M. (1985). Cognitive coping strategies and the problem of "inert knowledge." In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 2, pp. 65–80). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*, 4–16.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (4th ed., pp. 515–629). New York: Wiley.

Caccamise, D., & Snyder, L. (2005). Theory and pedagogical practices of text comprehension. *Topics in Language Disorders*, *25*, 5–20.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, *25*, 471–534.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, *19*, 237–248.

Cromley, J. G., & Azevedo, R. (2005). What do reading tutors do? A naturalistic study of more and less experienced tutors in reading. *Discourse Processes*, *40*, 83–113.

Cross, D. R., & Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of Educational Psychology*, *80*, 131–142.

Druckman, D., & Bjork, R. A. (1994). *Learning, remembering, believing: Enhancing human performance*. Washington, DC: National Academy Press.

Eckhoff, A. (2005, April). *Exploring the role of metacognition in an educational technology tool*. Paper presented at the meeting of the American Educational Research Association. American Educational Research Association, Montreal, Quebec, Canada.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363–406.

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive inquiry. *American Psychologist*, *34*, 906–911.

Franzke, M., Kintsch, E., Caccamise, D., Johnson, N., & Dooley, S. (2005). Summary Street: Computer support for comprehension and writing. *Journal of Educational Computing Research*, *33*, 53–80.

Gourgey, A. F. (2001). Metacognition in basic skills instruction. In H. J. Hartman (Ed.), *Metacognition in learning and instruction—Theory, research and practice* (pp. 17–32). Dordrecht, The Netherlands: Kluwer Academic.

Graesser, A. C., Bowers, C., Hacker, D. J., & Person, N. K. (1997). Anatomy of naturalistic tutoring. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches and issues* (pp. 145–184). Cambridge, MA: Brookline.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, *9*, 495–522.

Hacker, D. J. (1998). Definitions and empirical foundations. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 1–23). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hume, G., Michael, J., Rovick, A., & Evens, M. (1996). Hinting as a tactic in one-on-one tutoring. *Journal of the Learning Sciences*, *5*, 23–47.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kintsch, W., & Kintsch, E. (2005). Comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Current issues in reading comprehension and assessment* (pp. 71–92). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*, 79–97.

Landauer, T. K. (1998). Learning and representing verbal meaning: The latent semantic analysis theory. *Current Directions in Psychological Science*, *7*, 161–164.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction and representation of knowledge. *Psychological Review*, *104*, 211–240.

Landauer, T. K., Laham, D., & Foltz, P. (2003). Automatic essay assessment. *Assessment in Education*, *10*, 295–308.

Merrill, D. C., Reiser, B. J., Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *Journal of the Learning Sciences*, *2*, 277–305.

National Center for Education Statistics. (2004). *National assessment of educational progress. The nation's report card: Reading highlights 2003*. Washington, DC: U.S. Department of Education.

Scholastic (1999). *Scholastic Reading Inventory using the Lexile Framework technical manual forms A and B*. New York: Scholastic Inc.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic.

Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*, *22*, 333–362.

16

iSTART: A Web-Based Tutor That Teaches Self-Explanation and Metacognitive Reading Strategies

Danielle S. McNamara, Tenaha O'Reilly, and Michael Rowe
University of Memphis

Chutima Boonthum and Irwin Levinstein
Old Dominion University

iSTART (Interactive Strategy Training for Active Reading and Thinking) is a Web-based tutoring program that uses animated agents to teach reading strategies to young adolescent (Grades 8–12) and college-aged students. The program is based on a live intervention called Self-Explanation Reading Training (SERT) that teaches metacognitive reading strategies in the context of self-explanation. SERT was motivated by empirical findings that students who self-explain text develop a deeper understanding of the concepts covered in text, combined with a large body of research showing the importance of reading strategies such as comprehension monitoring, making inferences, and elaboration. SERT *was designed to improve self-explanation by teaching reading strategies and in turn to facilitate the learning of reading strategies in the context of self-explanation. SERT has been found to successfully improve students' comprehension and course performance at both the college and high school levels. iSTART was designed to deliver an automated version of SERT that could be more widely available and could adapt training to the needs of the student. This chapter reviews the literature that motivated SERT and iSTART, describes the iSTART program, and describes research that has demonstrated the effectiveness of the training programs. This research has shown that SERT is most beneficial for students with the least knowledge about the domain as well as the students who are less strategic or less skilled readers. In other words, iSTART is most beneficial to at-risk readers. Current efforts for iSTART center on expanding the types of strategy training in order to increase its adaptability to students' needs.*

The goal of this chapter is to describe an automated tutoring system called iSTART (*Interactive Strategy Trainer for Active Reading and Thinking),* which is designed to provide reading strategy instruction to adolescent students. The chapter is organized into six sections that broadly cover the history and future development of iSTART. In the first section, we provide the motivation for creating reading strategy interventions. Then we describe *Self-Explanation Reading Training* (SERT; McNamara, 2004b) and the empirical findings on the human-delivered strategy intervention that served as the impetus for creating iSTART. Next, we describe the basic characteristics of iSTART, including the overall structure of the program and the computational feedback system. We then describe our field studies on the effectives of the system in both college and high school settings and provide an analysis of the relationship between self-explanation quality and comprehension performance. In the section after that, we discuss the role of metacognition within the system and how metacognitive processes are encouraged implicitly through the structure of the program. Finally, we discuss the limitations of the current system and our future plans for improving iSTART and scaling up the program so that it can be easily incorporated into large school settings.

A NEED FOR READING STRATEGY INTERVENTIONS

The ability to comprehend written text is one of the most complex but critical activities people perform every day. From the time we wake up in the morning until the time we rest at night, we are bombarded with thousands of written messages, including advertisements, instructions, newspapers, magazines, and textbooks. Although successful comprehension can often appear effortless for skilled readers, the processes underlying the chain of activities necessary to comprehend text are complicated. Successful comprehension entails a highly integrated set of activities that involves both lower level decoding abilities (e.g., Perfetti, 1985; Shankweiler et al., 1999) and higher level integration abilities (Long, Oppy, & Seely, 1994; Magliano, Wiemer-Hastings, Millis, Muñoz, & McNamara, 2002; Oakhill & Yuill, 1996).

For example, in order to understand written text, individual letters must be decoded and combined to form words, sentences, and paragraphs. In addition, these combinations must adhere to a complex set of rules of grammar and syntax. Furthermore, the meaning must be derived from the integration of information contained in the text with the reader's prior knowledge (Kintsch, 1988, 1998). Comprehension can fail at any stage of the reading comprehension process. While there have been a variety of interventions to improve comprehension at the lower levels (Fuchs & Fuchs, 2005; Kuhn, 2005), our focus has been on improving higher level comprehension skills; that is, we have focused on students who can adequately decode but who are poor comprehenders (Cain, 1996; Cornoldi, DeBeni, & Pazzaglia, 1996; Hoover & Gough, 1990; Stothard & Hulme, 1996).

The need for reading comprehension interventions is clear from several sources. First, recent research has indicated that students in the United States typically score lower on measures of reading comprehension as compared to students in other countries. Even more startling is the statistic that as many as 37% of fourth-graders, and 26% of eighth-graders, cannot read at the basic level (National Assessment for Educational Progress, 2003). In other words, these fourth- and eighth-graders do not understand what they read. Reading comprehension difficulties are even more pronounced for minorities: over half of minority students cannot read at the basic level (National Assessment for Educational Progress, 2003). Second, other research has indicated that students rarely use reading strategies to help them comprehend text (Garner, 1990; Pressley & Ghatala, 1990; Pressley et al., 1992; Rothkopf, 1988) and, when they do use strategies, students often implement rudimentary and ineffective methods, such as repetition (Garner, 1990). Third, even when students read a text at the basic level, the level of comprehension is typically shallow and lacks the necessary depth for adequate understanding (Best, Rowe, Ozuru, & McNamara, 2005; Langer, 1989; Pressley et al., 1992). In short, there is a strong need for improving reading comprehension among students in the United States. Fortunately, interventions designed to improve comprehension have been successful (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, De Leeuw, Chiu, & LaVancher, 1994; Johnson-Glenberg, 2000;

McNamara, 2004b; Meyer et al., 2002; Palincsar & Brown, 1984; Paris, Cross, & Lipson, 1984; Pressley et al., 1992).

SERT: SELF-EXPLANATION READING TRAINING

In response to the growing need for reading strategy training, McNamara and her colleagues (McNamara, 2004b; McNamara & Scott, 1999) developed a reading strategy training program called SERT (Self-Explanation Reading Training). SERT is a human-delivered reading strategy intervention that is based on nearly 30 years of research and theory on learning, memory, and reading comprehension. The broad scope of the training focuses on teaching students to become more active in constructing meaning through an integrative process of building a coherent model of the text in relation to the learner's prior knowledge. Particular focus has been directed at the active production of knowledge as opposed to the passive reception of concepts within the text.

The positive effects of active production on learning have been supported from the literature on learning and memory (Healy et al., 1993). For example, research on the generation effect shows that information produced by oneself is better remembered than information read more passively (e.g., McNamara & Healy, 2000). In general, research has shown that learning is improved when learners are forced to make more inferences and link new information with prior knowledge. SERT capitalizes on that notion in the sense that it encourages the reader to more actively approach text and use whatever knowledge he or she has available to make sense of it.

The backbone of SERT was largely motivated by research on self-explanation (Chi et al., 1994). The central idea is that students who explain the meaning of a text are more likely to make inferences, solve problems, construct coherent mental models, and develop a deep understanding of the information in the text (Chi et al., 1989, 1994). In short, self-explanation improves learning. However, not all readers successfully self-explain text. Thus, SERT builds on the benefits of self-explanation by providing a comprehensive and detailed training program that incorporates self-explanation, metacognitive skills, and reading comprehension strategies.

SERT is broken up into three core training sections: (a) introduction, (b) demonstration, and (c) practice. In the introduction, students are introduced to the five core reading strategies and provided with examples of how these strategies could be applied while reading texts. The first strategy is *comprehension monitoring*, which serves as an executive manager of the students' overall learning process. Students are taught to monitor whether they understand what they are reading. If students reach an impasse in their understanding, they are encouraged to use the other strategies to rectify the comprehension failure.

The second strategy is called *paraphrasing* and is used as a catalyst for self-explanation. Describing the text in one's own words serves two functions. First, it allows the reader to transform the material into a representation that is more familiar and consequently more memorable. Second, the ability to paraphrase roughly translates into the most basic level of comprehension because, to paraphrase successfully, one must be able to process the basic structure and relations of the sentence to transform the verbatim text into more familiar words.

The third strategy, called *elaboration*, builds on paraphrasing by encouraging readers to go beyond the text by using their prior knowledge, common sense, and logic to elaborate the text. Empirical findings indicate that prior knowledge plays an important role in learning (Shapiro, 2004; Thompson & Zamboanga, 2003). In particular, prior knowledge is critical in helping readers make the necessary inferences to fill gaps within the text (McNamara, Kintsch, Songer, & Kintsch, 1996). However, even when readers do not have any specific knowledge related to a text, SERT encourages them to elaborate by using general knowledge, logic, reasoning, or common sense to repair gaps in their understanding. In this manner, SERT encourages active, repair-directed processing, as opposed to passivity, when readers encounter an impasse.

The fourth strategy is *prediction*. Although research indicates that predictions are relatively infrequent (Magliano, Trabasso, & Graesser, 1999), teaching students to predict what will occur next can be a useful exercise in metacognition. Forming a prediction requires the reader to make plausible guesses about the future text content based on the current available evidence. Most important, checking to see whether a prediction is validated serves as a form of self-regulated learning.

The fifth strategy taught in SERT is called *bridging*. Bridging teaches the reader to link the concepts within the various parts of the text. Bridging is critical, because many texts are not written in a manner that explicitly maps how the various concepts within the text are related (Beck, McKeown, & Gromoll, 1989). Bridging allows the reader to link concepts in both the proximal and distal sentences of the text to form a more global model of the content. In other words, making bridges between the elements of the text fosters many of the inferences necessary to successfully comprehend the material.

These strategies can be mapped onto levels of comprehension assumed by theories of text comprehension (see chap. 1, this volume; Kintsch, 1998). Strategies such as paraphrasing and bridging help the reader to better understand the basic meaning of the text and thus strengthen the reader's textbase situation model level of understanding. In turn, the last three strategies (elaboration, prediction, and bridging) strengthen the reader's (textbase) situation model level of understanding. The principal point conveyed to the students is that they need to understand not only concepts within the text, but also relationships between concepts in the text and relationships between the text and what they already know. The inference-generation strategies combined with the process of self-explanation help the reader to form a more coherent situation model. Making inferences is critical to successful comprehension because inferences help the reader to construct a more coherent mental representation of the text. In essence, the goal is for the reader to seek coherence by using both the text and prior knowledge to create links between concepts.

Once the students have completed the introduction phase of the training, they proceed to the demonstration phase. During this phase of SERT, participants watch a video depicting a student reading and self-explaining a text about forest fires. Participants refer to an accompanying transcript while viewing the video. The video is paused at various points, and participants identify and discuss the strategies being used by the student in the video. During the final phase, practice, the participants work in pairs to practice self-explanation while reading a chapter from their science textbook. The participants take turns self-explaining, alternating after each paragraph. At the end of each paragraph, the partner who listens (and is not self-explaining) summarizes the paragraph.

Overall, SERT can be taught to small groups of students in about 3 hours of training. Empirical studies on the effectiveness of SERT have been very promising. Our results have shown that SERT is more effective than controls in improving college students' comprehension (Magliano et al., 2005, Experiment 1; McNamara, 2004b) and science course performance (McNamara, 2006; for reviews, see McNamara, 2004a; McNamara, Best, & Castellano, 2004; McNamara & Shapiro, 2005). SERT has also improved comprehension for high school students when compared with two other strategy interventions that emphasized knowledge activation and comprehension monitoring (O'Reilly, Best, & McNamara, 2004).

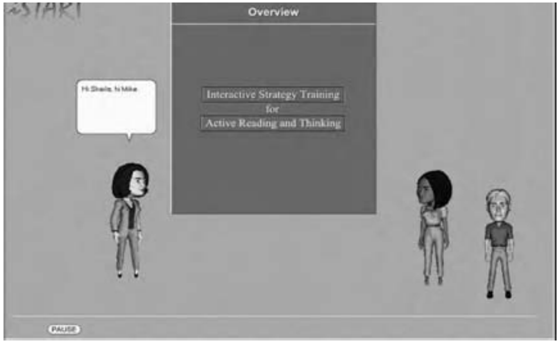
Of particular importance is the finding across studies that the effects of SERT are most evident for the students who show the lowest comprehension that is, those with either low domain knowledge or low reading skills (McNamara, 2004b; McNamara, O'Reilly, Best, & Ozuru, submitted; O'Reilly, Best, & McNamara, 2004; O'Reilly, Sinclair, & McNamara, 2004a). This result indicates that the training is effective for students who need it the most: those who do not possess enough knowledge or those who do not automatically understand the relationships between concepts in the text. Protocol analyses have further revealed that SERT helps students compensate for their low knowledge by elaborating the text with their general knowledge by using logic and common sense to make the inferences to bridge knowledge gaps. In short, SERT has been successful in improving students' text comprehension, and the effects of the training seem to be more evident for those who need it most.

One question that this result raises concerns the relationship between incorrect explanation and comprehension. One might assume that low-knowledge and less skilled readers would be more likely to produce explanations with inaccurate information. In contrast, we have not observed more inaccuracies in explanations for either low-knowledge or less skilled readers after they have been provided with strategy training (McNamara, 2004b; Ozuru, Best, & McNamara, 2004). Moreover, McNamara (2004b) found a positive relationship between inaccurate elaborations and comprehension and no relationship between inaccurate bridging inferences and comprehension. Thus, elaboration helps to improve comprehension, regardless of whether the elaboration is accurate. Hence, making inferences is critical to successful, coherent comprehension. Whether these inferences are initially accurate is not a driving factor, at least not in the context of self-explanation.

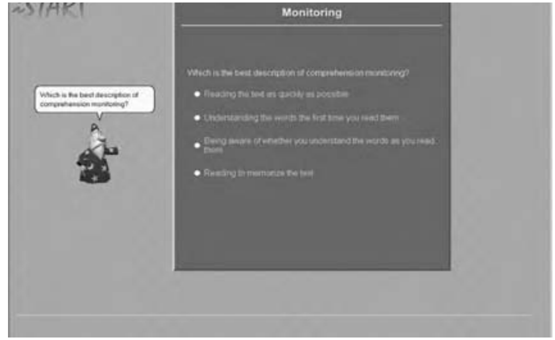
iSTART: THE PROGRAM

Although the effects SERT have been promising, there are several drawbacks to the human-delivered method of training. First, human training is costly because of the amount of time and resources required to train people how to teach SERT. Second, human-delivered feedback is inconsistent, and when given in group format it cannot be easily adapted to the individual participant. Third, human training is not accessible to a large number of people at any time of the day. Given these constraints, McNamara, Levinstein, and Boonthum (2004) developed an automated version of SERT called iSTART (Interactive Strategy Trainer for Active Reading and Thinking). iSTART solves many of the problems associated with human-delivered training because the program can be adapted to the individual needs of the reader and, given that the system is web-based, the training can be scaled up to serve large-scale needs.

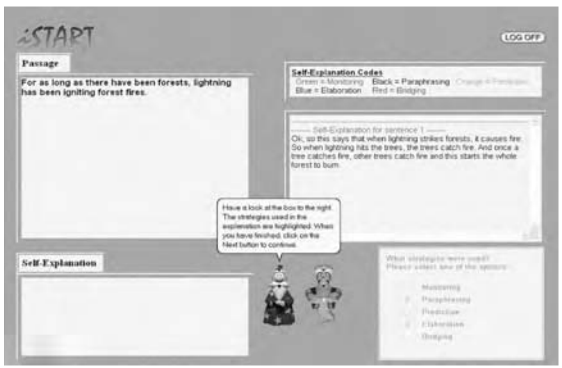
iSTART, like SERT, is composed of three sections: (a) introduction, (b) demonstration, and (c) practice (see Fig. 16.1). Each section provides progressively more interaction with the trainee in regard to reading strategy use while self-explaining text. As stated earlier, the purpose of the iSTART trainer is to provide readers with strategies to comprehend texts at a deep level. iSTART provides the trainee with these abilities by teaching reading strategies in a scaffolded, structured manner. Each section of the program is hosted by animated pedagogical agents that provide the trainee with guidance and instruction using generated speech and gestures. At first, the agents provide self-explanations while the trainee watches, but as the trainee progresses through the modules he or she creates self-explanations that are evaluated by the agents.



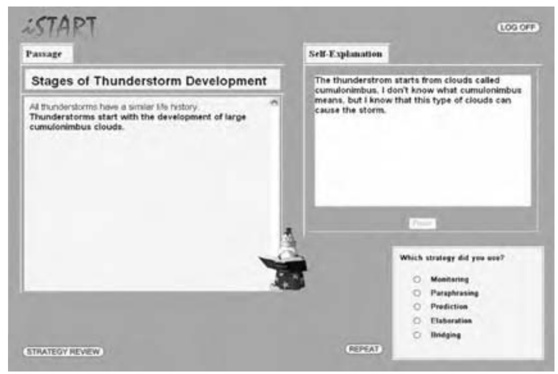
A. Screen shot of the three pedagogical agents (Dr. Julie, Sheila, and Mike) who deliver training during the Introduction Module of iSTART.



B. Screen shot of Merlin who gives a quiz during the Introduction Module of iSTART.



C. Screen shot of Merlin and Genie who deliver training during the Demonstration Module of iSTART.



D. Screen shot of Merlin who delivers training during the Practice Module of iSTART.

Figure 16.1. iSTART sample screen shots.

The introduction module introduces trainees to the reading strategies within a vicarious learning environment (McNamara, Levinstein, & Boonthum, 2004). The introduction is hosted by three pedagogical agents: (a) Dr. Julie (the instructor agent), (b) Mike, and (c) Sheila (two learner agents). During the introduction, Dr. Julie introduces reading strategies and instructs Mike and Sheila on how to use the strategies. At this point, the learning environment is primarily vicarious: The trainee observes proper use of the strategies from Mike and Sheila but does not participate in their use. After the reading strategy is introduced by Dr. Julie and demonstrated by Mike and Sheila, the trainee must answer some review questions pertaining to the definition and to the application of the strategies used in example sentences.

The demonstration module provides an illustration of self-explanation through an interaction between two new pedagogical agents: (a) Merlin (the teacher) and (b) Genie (the learner), who use reading strategies to explain sentences in a science text. Merlin guides Genie in the use of self-explanation and reading strategies, and Genie produces self-explanations, typically using a combination of strategies. After Genie self-explains a sentence, Merlin assesses the quality of the self-explanation, thus providing an example of the feedback a trainee might receive when self-explaining in the subsequent, practice module. The trainee is questioned by Merlin each time that Genie produces an acceptable self-explanation. For example, the trainee might be asked what type of strategy Genie used, or where a particular strategy can be found within a self-explanation.

During the demonstration module, Merlin's questions adapt to the level of the trainee, providing different levels of support. Merlin has four levels of question style. The most supportive question style, Level 1, states and defines the strategy Genie used, rather than letting the trainee guess. The trainee is then asked to indicate where in Genie's self-explanation that strategy was used. Level 2 questions are identical to Level 1, but the definition is removed. Level 3 questions ask the trainee what strategy Genie used, and the trainee is presented with a menu from which to choose. Two different strategies are usually required to be identified by the trainee in each self-explanation. With each strategy, follow-up questions are asked to focus the trainee on the details of the strategy and its use (e.g., "Click on the part of the text to which this explanation linked"). The trainee begins the demonstration module with questions at Level 3. Level 4 questions are the same as Level 3 but do not contain follow-up questions. Merlin will progress to less structured questions if the trainee has a high success rate, whereas Merlin will provide structured, focused questions if the success rate is low.

During the practice module, the trainee practices the reading strategies he or she has learned. Merlin instructs the trainee to self-explain specific sentences in a text and the trainee's self-explanations are rated by the iSTART system so that Merlin can provide feedback. Merlin's feedback, based on a computational linguistic algorithm, either asks for more information in the self-explanation or praises the self-explanation and allows the trainee to continue (McNamara, Boonthum, Levinstein, & Millis, in press; Millis et al., 2004; Millis, Magliano, Wiemer-Hastings, Todaro, McNamara, in press). If allowed to proceed, the trainee may be asked what strategy was used or specifics about the strategy, such as where in the self-explanation it is located.

Merlin's feedback is based on both the self-explanation and the target sentence, or the sentence being explained. To account for misspellings, words in the self-explanations are matched using a soundex (Knuth, 1998) transformation that drops vowels and maps similar characters to a single character. The appropriateness of the trainee's self-explanations is first assessed in three ways: (a) length, (b) similarity, and (c) relevance. Self-explanations that have little in common with the target sentence and self-explanations that are simple restatements of the target sentence are unacceptable. The self-explanation must be sufficiently different from the target sentence but still relevant to it. If the self-explanation is not long enough, different enough, or sufficiently relevant, more information is requested of the trainee. For example, if the self-explanation is not long enough (compared with the length of the target sentence), Merlin asks for more information to be added to the self-explanation.

Relevance and similarity of the self-explanation and target sentence are established using content words from the target sentence and the other sentences in the text. The proportion of content words that overlap with the target sentence generally corresponds to the detection of a restatement or paraphrase of the sentence. The overlap in content words in the self-explanation with the content words from the other sentences in the text generally corresponds to a detection of the relevance of the self-explanation. Relevant self-explanations must contain a certain number of content words or associates. However, if the self-explanation contains too many content words from the target sentence in proportion to other content words, then the self-explanation is considered too similar. The feedback system in that case would ask for more information from the trainee.

If the self-explanation passes the initial screening, it is evaluated with respect to quality. The self-explanation quality is based on the same three factors as the initial assessment of the sentence but in addition uses latent semantic analysis to judge a general conceptual overlap between the self-explanation and the text. The quality judgment guides the final feedback that Merlin provides to the trainee. For example, if the quality is low, Merlin might suggest making a bridging inference next time, or, if the quality is high, Merlin would tell the trainee that the self-explanation was excellent.

EVIDENCE THAT ISTART WORKS

Empirical studies on the effectiveness of iSTART have been encouraging. Studies at both the college and high school levels have indicated that iSTART improves text comprehension and strategy use over control groups. There are two essential questions that the iSTART research program has addressed: (a) Is iSTART as effective as SERT? and (b) Who benefits from iSTART?

iSTART Versus SERT

One of the initial questions after developing our first version of iSTART was whether it matched SERT (the live classroom training) in effectiveness. In separate studies, Magliano et al. (2005) showed that both SERT and iSTART were effective, but they did not directly compare the two training conditions. O'Reilly, Sinclair, and McNamara (2004b) conducted a study with college students enrolled in an introductory biology course to examine whether iSTART resulted in comprehension gains comparable to SERT. Although automating the SERT intervention has several advantages, one potential problem is that automation may influence the effectiveness of SERT. In this study, the course laboratory sections were randomly assigned to one of three conditions: (a) live SERT (trained by a human instructor), (b) iSTART (trained by the computer program), and (c) a control condition with no training (students instead read a text and answered questions concerning it). After training, participants read a passage on cellular mitosis (see McNamara, 2001). As shown in Figure 16.2, we confirmed that iSTART and SERT students answered more questions correctly than did students in the control condition. However, this main effect of condition was qualified by a significant interaction between condition and question type. There were two types of questions: (a) text-based and (b) bridging inference. Text-based questions could be answered on the basis of individual sentences from the text, whereas bridging questions required multiple sentences and understanding the relationship between them. Text-based questions are intended to assess the readers' basic understanding of the text content, whereas bridging questions are intended to assess deeper level understanding that results from greater inferencing while reading. The interaction of condition and question type in this study indicated that the SERT and iSTART participants outperformed control participants on text-based questions, but not bridging-inference questions. The effect of training was marginal for the bridging questions. Thus, both SERT and iSTART improved students' comprehension, particularly at a basic, textbase level of understanding.

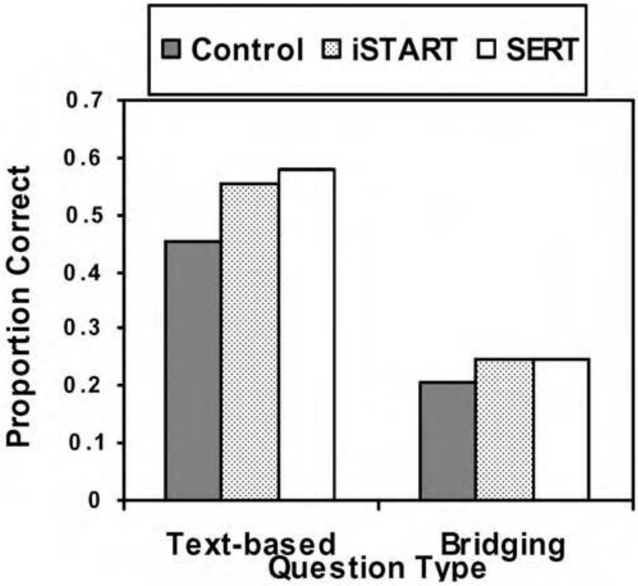


Figure 16.2 Comprehension of cell mitosis text as a function of training condition and question type. iSTART = Interactive Strategy Training for Active Reading and Thinking; SERT = Self-Explanation Reading Training.

This result is somewhat disturbing, because the goal of iSTART is to improve deep-level understanding of the text. However, the results are congruent with three other studies that used the same cellular mitosis text (McNamara, 2001, 2004b; O'Reilly & McNamara, 2006). In all three studies, the effects of the manipulation intended to induce active processing produced effects on the text-based questions but not the bridging-inference questions. In McNamara's (2001) study and in O'Reilly and McNamara's (2006) study, the advantage of inducing inference generation for high-knowledge readers through less cohesive text appeared only on the text-based questions, and in McNamara's (2004b) study the advantage of SERT appeared only on the text-based questions. One explanation of these results simply relates to the high difficulty of the text and the topic of the text. That is, we might expect the level of a text's difficulty to influence whether improvement occurs at a lower or deeper level of processing: Improvement at deeper levels of processing may be more likely for texts with moderate difficulty, and improvements at lower levels of processing may be more likely for texts with high difficulty.

This hypothesis is related to the notion of the *zone of proximal development* (Vygotsky, 1978), the idea that scaffolding helps a learner achieve a level of learning that is otherwise unachievable without scaffolding. Comprehension of the cellular mitosis text without scaffolding is very poor—virtually a complete lack of understanding. With scaffolding, across these experiments, the readers were able to understand the basic ideas in the text. They were not, however, able to generate inferences and make links between ideas to allow them to perform well on the bridging inference questions.

So, we see that the idea that traces of deep processing will appear only on bridging inference questions may be too simplistic. In some cases, the benefits of deeper processing may lead only to a basic understanding of the information in the text. The important comparison is whether it is the case that, without that deep or active processing, the reader would understand very little of the text.

Who Benefits From iSTART?

There have been two intertwined questions regarding who would most benefit from iSTART. The first question was whether iSTART would benefit high school students. Our research has shown that both SERT and iSTART benefit college students (McNamara, 2004b; O'Reilly et al., 2004b) and that SERT was beneficial to high school students (O'Reilly, Best, & McNamara, 2004; O'Reilly et al., 2004a). One of our main goals was for iSTART to be used in high school classrooms, and thus a central aim has been to investigate its effectiveness with that population. Much of the research we have conducted with high school students is ongoing, in the sense that it has not yet been published. Thus far, five classroom studies in 35 high school classrooms have investigated the effectiveness of iSTART. This research has included almost 1,000 students. However, this research is ongoing because the complexity of scoring and analyzing the vast amount of data. Thus, in this chapter we discuss only our published research, leaving a description of our high school classroom data for a future venue. Nonetheless, we can say that the preliminary analyses of our data are very promising, indicating that iSTART is quite effective with high school students and that in some cases raises low-comprehension students up to par with regular students.

A second set of questions has regarded whether the effects of iSTART depend on individual differences and how these effects manifest. Are there certain readers who do not need iSTART, or for whom it is too challenging? What is the role of prior domain knowledge? Do the benefits of iSTART depend on reading skill? There are two central goals to answering these types of questions. The first goal is to discover if there are certain students who do not need training, and the second goal is to discover if certain students need a different type of training.

One study investigated the effect of iSTART on adolescent students' comprehension and strategy use (McNamara et al., in press; O'Reilly, Sinclair, & McNamara, 2004a). This study also examined whether the students' prior knowledge of reading strategies interacted with the benefits of strategy training (McNamara et al., in press). Half of the students were provided with iSTART, and the students in the control condition were given a brief demonstration on how to self-explain text. All of the students then self-explained a text about heart disease and answered text-based and bridging-inference questions. We found that both iSTART training and prior knowledge of reading strategies significantly improved the quality of self-explanations and, in turn, comprehension. In addition, we saw that the benefits of reading strategy instruction depended on prior reading strategy knowledge. For low strategy knowledge participants, the effects of iSTART were more pronounced at the more literal text-based level. Conversely, for high strategy knowledge students, the effects of iSTART were evident on more difficult and integrative bridging-inference questions. Protocol analyses indicated that iSTART improved the quality of the students' self-explanations and, in turn, the quality of the students' explanations was indicative of improved comprehension scores.

Thus, it appears that the majority of the students benefited from iSTART, but in different ways, and according to their zone of proximal development. Students with less knowledge of reading strategies needed to learn how to develop a coherent understanding of the basic information conveyed in each sentence of the text. In contrast, those with more prior knowledge of reading strategies were able to make more bridging inferences and elaborations, which translated to better performance on the bridging inference questions.

Magliano et al. (2005) found a similar pattern of results when they investigated whether and how the benefits of iSTART depended on the students' prior reading skill. In their study, college students read and self-explained two science texts before and after iSTART training. After reading the two texts, the students responded to eight short-answer comprehension questions that corresponded to each text. Their reading skill was measured with the Nelson Denny Comprehension Test (Brown, Fishco, & Hanna, 1993). The experimenters found that skilled readers answered more bridging questions correctly after training, whereas less skilled readers improved on the text-based questions. Thus, more skilled readers learned strategies that allowed them to make more connections within the text, and this ability was most apparent on the bridging inference questions. In contrast, the less skilled readers learned the more basic-level strategies (e.g., paraphrasing) that allowed them to make sense of the individual sentences. Future research will reveal whether additional, extended training will help less skilled or less strategic readers to go beyond sentence-level understanding and develop the skills necessary for a coherent, global understanding of challenging text.

These results indicate that the students will make progress in their area of proximal development (e.g., Vygotsky, 1978). Readers first need to learn to form an adequate representation of the text-based information—in essence, the information presented in each individual sentence. Then, readers can learn how to understand the text at a deeper level by processing the relationships between the ideas conveyed across sentences and making links to world knowledge. iSTART can allow this progressive improvement by providing training at various levels of processing.

THE ROLES OF METACOGNITION, ENGAGEMENT, AND REFLECTION IN iSTART

One important aspect of iSTART is that the skills it teaches are essentially metacognitive in nature. *Metacognition* refers to an individual's monitoring of cognitive processes and knowledge and use of cognitive processes for successful learning. When applied to reading, metacognition involves the reader's monitoring of whether the written material is successfully comprehended, coupled with active reading strategies that enhance and repair comprehension. iSTART provides metacognitive training first by explicitly teaching the students about metacognitive reading strategies and how to use them. Additionally, iSTART adapts the constructivist modeling–scaffolding–fading paradigm to an automated curriculum by carefully increasing the engagement of the students and turning their attention to their own thought processes.

Metacognition and Engagement

The iSTART program initially provides the students with a safe, low-stakes environment in which they play the role of the observer and gradually modifies that role until the students become intellectually involved in self-explaining new texts. iSTART directs the student through three curricular phases: (a) introduction, (b) demonstration or modeling, and (c) practice. There are four types of interaction in the iSTART curriculum, listed here from the least demanding to most demanding for the student:

1. *Navigational clicks*. The student clicks on a button to progress to the next step or to repeat an example.

2. *Well-defined choice clicks*. The student selects an answer from clearly defined items; for example, a set of radio buttons.
3. *Fuzzy choice clicks*. The student makes a choice, but the choices are not clearly defined; for example, identifying where a certain strategy is used within an explanation.
4. *Production of text in response to a prompt*. The student is asked to type an explanation of a given sentence.

The ultimate aim of iSTART goes beyond this last item: the production of unprompted explanation directed by cognitive monitoring. As students progress through this curriculum, the changes in engagement are evident in the changes in the type of interaction demanded by the program: the proportion of navigational and well-defined choice clicks declines, and fuzzy choices and the production of text increase. This progress is scaffolded by a simulated social environment consisting of animated characters that provide instruction, examples, coaching, and assistance. In the next few sections, we describe how metacognition and engagement emerge within the three modules of iSTART.

Introduction. The introduction is presented by three animated characters, as described earlier. Dr. Julie presents each reading strategy with definitions and examples to the animated students (Mike and Sheila), who ask questions and try out the strategies under her coaching. The actual student who is being tutored merely observes this interchange and participates only by making limited choices: clicking on a button to continue after having studied an example on the blackboard or clicking on a button to repeat an example or to see a second example. The student's interaction at this point is minimal, consisting of a few navigational clicks in each of the eight modules.

The level of engagement is then stepped up, with brief quizzes that follow the presentation of each strategy. The quiz is actually a pedagogical tool for providing additional instruction when needed as well as a means of assessing the learner. Each quiz consists of four multiple-choice questions that tap the student's understanding of the definition of the strategy and the ability to identify good examples of the strategy. Although the quizzes put the student "on the spot" much more so than the animated expository modules, they also scaffold the student by providing prompts and hints when an incorrect answer is chosen. In the first instance of an incorrect answer choice, the student is asked to try again. With a second error, the student is reminded of the definition of the strategy. If the student makes any errors, the final, correct choice is followed by an explanation of the correct answer.

Demonstration. Whereas the introduction alternates several minutes of observation with several minutes of testing, the demonstration module shortens the period of this alternation and makes more difficult demands on the student. Merlin, the coach seen earlier during the quizzes, supervises a new character: Genie, a surrogate learner who reads and self-explains a passage sentence by sentence. Genie's explanation remains available to the student in a text box on the screen. While Genie appears to be the one on the spot to produce a self-explanation acceptable to Merlin, the student, who observes their interaction, is also on the spot, because she or he is questioned about the strategies used in each self-explanation. Although they are questioned more often during demonstration than the introduction, students are still in a protective environment in the demonstration section because Merlin gives them several chances to answer and because the trainer adapts the mode of questioning to the students' success (as described earlier). At a lower level of questioning, Merlin presents the students with a list of strategies and asks them to click on any strategy used in the self-explanation. More challenging follow-up questions require the student to identify where a certain strategy appeared within the self-explanation or what text is related to Genie's self-explanation (i.e., to identify the source of a bridging inference). These are interactions of the fuzzy choice variety. Answering these questions demands more difficult thought processes than answering the questions in the introduction, in which the students had to decide among well-defined choices, such as which of several explanations counted as bridging. Now, Genie's explanation must be mentally separated into parts and the parts matched up with one of the strategies on the list. When students have difficulty with this mode of questioning, the trainer adapts by simplifying the question.

Overall, the demonstration module requires more interaction and higher level thought processes from the students than required during the introduction. No longer is the interaction primarily among the animated characters while the trainee mostly navigates through the material. Instead, the trainees alternately observe and participate by analyzing the competencies they will be expected to exhibit in the next phase of training. In the introduction, all clicks are made on constrained choices, whereas in the demonstration section more than half are fuzzy choices made on the text of an explanation or the passage being explained, a process that requires careful reading or parsing.

Practice. In the practice module, the student takes on the role that Genie played in the demonstration module and types explanations of the sentences in a new text, under Merlin's coaching. Here, the student is much more on the spot and is no longer observing others but fully participating by creating explanations.

Merlin coaches the student by encouraging him or her to develop an explanation of at least minimal acceptability and then provides more or less enthusiastic feedback depending on the quality of the explanation. Merlin also asks the student to identify the strategies used in the explanations just as he asked about Genie's explanations during the demonstration module. The student's engagement with the program in the practice module goes beyond what it was in the demonstration, a combination of well-defined and fuzzy choices because they produce text at Merlin's prompts and analyze their own explanations rather than Genie's. The learner no longer simply clicks to interact with the program but focuses on explaining the sentences as they are presented.

Modeling–Scaffolding–Fading

Modeling and Scaffolding. iSTART makes sophisticated use of the modeling–scaffolding–fading paradigm. Use of the strategies to self-explain a sentence is modeled in the first two parts of the curriculum. Scaffolding is provided throughout but fades as the student becomes successful. In the introduction, the use of strategies is modeled by Dr. Julie, who provides many examples of good explanations, and by the student characters, Mike and Sheila, who are coached through the process as well. This modeling is reinforced because the explanations are heard, seen in bubbles as the words are spoken, and preserved on a blackboard for later discussion.

In the demonstration module, Genie, although cast as a student, is actually an expert self-explainer most of the time (early on, Genie produces an inadequate explanation so that the students can observe the sort of coaching they will receive when Merlin complains about their explanations in the practice module). His explanations typically use several strategies and are generally longer than the explanations actually provided by students using iSTART. As in the introduction, the explanations are presented verbally, as text in the balloons as they are spoken, and are preserved in a text box on the screen. Modeling can be effective only if the student attends to it. In this module, the student is strongly encouraged to attend to the modeling by being questioned in detail about Genie's explanations. Even if the student fails to pay attention while Genie is speaking, the explanation is reproduced on the screen while the student is being questioned.

The trainer provides the students with a great deal of scaffolding as they make their observations and analyses. In the introduction module, the students are able to repeat the examples that model self-explanation at the click of a button, and they are coached in the quizzes. In the demonstration module, Merlin matches his style of questioning to the success of the student, and a color-coded analysis of each self-explanation is provided for study before the student proceeds to the next sentence. The practice module continues the scaffolding. Merlin identifies explanations that are irrelevant, too short, or too similar to the original sentence and encourages the student to modify and expand them. When a student seems to be having too much trouble, the Genie character appears, coming to the rescue and offering more detailed assistance. Because students are regularly asked which of the strategies they have used and are shown a list from which to choose, they are constantly reminded of all the available strategies. When a student persists in identifying only comprehension monitoring and paraphrasing, Merlin encourages him or her to use other strategies as well. Finally, the student is motivated to construct better self-explanations by means of Merlin's feedback, which can range from ho-hum to enthusiastic.

Fading. An essential part of the paradigm is that students learn to proceed without scaffolding so that they can use their competencies autonomously. iSTART brings fading into play after the introduction module by reducing scaffolding in line with student success. In the demonstration module, this takes two forms. As a student becomes more successful in identifying the strategies used, the mode of questioning becomes less supportive. In addition, with continued success fewer follow-up questions are posed, allowing the student to move more quickly through the curriculum. In the practice module, a similar tactic is used: Successful students are asked fewer follow-up questions and queried less about the strategies they used. Merlin may just compliment their explanation and move on to the next sentence. The initial coached sessions of practice may be followed in subsequent weeks with refresher practices that are almost entirely uncoached unless the student runs into difficulty.

Reflection

The competencies taught by iSTART are unusual in that no specific subject-matter mastery is involved. Indeed, although the trainer seems to place a high value on identifying strategies used in explanations, that skill is promoted only as part of a pedagogy that ultimately engages the students in the practice of generating effective self-explanations. Strategy identification is a mechanism by which the students learn to think about and direct their own thinking as they read, that is, as they learn to read reflectively. iSTART induces this change by building a bridge between the concepts of strategic reading and the performance of strategic reading. In the introduction, the concepts are introduced and then modeled individually by Dr. Julie. Mike and Sheila then model learning to use self-explanation. The students observe all of these, encountering self-explanations in various forms. They also identify types of self-explanations during the quizzes. This identification practice is put to work in the demonstration module, where the students observe Genie self-explaining, again encountering the explanations as both performance and text, and practice the identification skills on Genie's explanations. Finally, in the practice module the students produce their own self-explanations. Here, the fact that the explanations are typed is crucial. There is virtually no difference between applying strategy identification to Genie's explanations and their own. By varying the students' perspective from that of observer, to critic, to producer of self-explanations, the students learn autonomous, reflective self-explanation.

THE FUTURE OF iSTART

Our current efforts are being directed to expanding the iSTART program so that it is more adaptive to student needs and can be more easily used in a classroom. The results discussed earlier in the chapter showed that for both adolescents (McNamara et al., in press) and college students (Magliano et al., 2005), the benefits of iSTART for less strategic or less skilled readers occur only at the textbase level of comprehension. These results suggest that less skilled readers may benefit from more extensive training than currently provided by iSTART. More extensive training could allow less skilled readers the necessary time and practice to be able to successfully use the more complex strategies, such as bridging inferences and elaborations. Some of our current efforts are directed toward expanding iSTART so that it provides more extensive training and with a greater range of strategies. Developing procedures to tailor feedback to meet the needs of high-risk, struggling readers should make iSTART more effective for a wider range of students. For example, assessing prior knowledge of reading strategies will allow us to better tailor the training. In future versions of iSTART, less skilled and low-strategy knowledge students will receive more training overall. First, they will receive more training in lower level strategies, and more positive feedback for strategies such as paraphrasing. This will then provide them with a stronger foothold to help them to move on to deeper level strategies. In contrast, we will continue to push more skilled students to go beyond the text by using strategies such as elaboration to create coherence.

We are also increasing the number and variety of texts that can be self-explained. In this way, the students will have the opportunity to experience a greater amount of training and to learn the strategies with a wider variety of text genres. A greater amount of training will allow less skilled students to develop the higher level skills necessary to build a deep understanding of difficult texts. The use of a wider variety of text genres will help students learn when and how to use self-explanation and other reading strategies in multiple contexts.

Finally, we are developing a teacher interface to help manage what the students read and to monitor their progress. The interface will allow teachers to more easily make use of iSTART in their classrooms according to their own needs. Incorporating a computerized trainer into a classroom is not as simple as just giving it to the teacher and expecting that it be used consistently or successfully.

Many teacher needs must be met to reach our goals. First, the teachers must understand the need for reading strategy training and be receptive to intelligent tutoring systems. Hence, we are developing an automated program for teachers to provide information about the importance of reading strategies and information about iSTART, particularly how it helps students to read and understand difficult material. Second, the program must be easy for the teacher to use and answer the questions that the teacher has concerning the program and the students' progress. Therefore, a program is being developed to facilitate teachers' use of iSTART in the classroom. Third, teachers are pressed to cover an increasing amount of content, causing many to drop the deep-learning activities that take extra time. Thus, the learning material covered in iSTART must be relevant to the course demands. Currently, iSTART tutors students using a limited number of texts, covering content that may or may not be relevant to particular courses. Therefore, we are increasing the number of course topics that can be covered during demonstration and practice and increase the difficulty range of practice texts. Teachers will be able to assign texts or topics for self-explanation training that are being covered in class—reducing time taken away from course material and thereby improving students' understanding of the course topic.

Like many of the other reading strategy interventions presented in this volume, our motivation for the iSTART project is to address what we view as a critical need in our educational system—to provide a large number of students with reading strategy training based on empirically supported and theoretically grounded reading strategy research. Our research has clearly documented the success of SERT and iSTART in terms of improving self-explanation abilities, comprehension success, and course grades. Hence, we are confident that iSTART has the potential to have a marked impact on students' reading abilities across the United States.

REFERENCES

Beck, I., McKeown, M., & Gromoll, E. (1989). Learning from social studies texts. *Cognition and Instruction*, *6*, 99–158.

Best, R. M., Rowe, M. P., Ozuru, Y., & McNamara, D. S. (2005). Deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders*, *25*, 65–83.

Brown, J., Fishco, V., & Hanna, G. (1993). *Manual for scoring and interpretation, Forms G & H*. Chicago: Riverside.

Cain, K. (1996). Story knowledge and comprehension skill. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 167–192). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145–182.

Chi, M. T. H., De Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self explanations improves understanding. *Cognitive Science*, *18*, 439–477.

Cornoldi, C., DeBeni, R., & Pazzaglia, F. (1996). Profiles of reading comprehension difficulties: An analysis of single cases. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 113–136). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Fuchs, D., & Fuchs, L. (2005). Peer-Assisted Learning Strategies: Promoting word recognition, fluency, and reading comprehension in young children. *Journal of Special Education*, *39*, 34–44.

Garner, R. (1990). When children and adults do not use learning strategies: Toward a theory of settings. *Review of Educational Psychology*, *60*, 517–529.

Healy, A. F., Clawson, D. M., McNamara, D. S., Marmie, W. R., Schneider, V. I., Rickard, T. C., et al. (1993). The long term retention of knowledge and skills. In D. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 30, pp. 135–164). New York: Academic.

Hoover, W., & Gough, P. (1990). The simple view of reading. *Reading and Writing*, *2*, 127–160.

Johnson-Glenberg, M. (2000). Training reading comprehension in adequate decoders/poor comprehenders: Verbal versus visual strategies. *Journal of Educational Psychology*, *92*, 772–782.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.

Knuth, D. E. (1998). *The art of computer programming: Vol. 3. Sorting and searching* (3rd ed.). Reading, MA: Addison-Wesley.

Kuhn, M. (2005). A comparative study of small group fluency instruction. *Reading Psychology*, *26*, 127–146.

Langer, E. J. (1989). *Mindfulness*. Reading, MA: Addison-Wesley.

Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1456–1470.

Magliano, J. P., Todaro, S., Millis, K. K., Wiemer-Hastings, K., Kim, H. J., & McNamara, D. S. (2005). Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research*, *32*, 185–208.

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91*, 615–629.

Magliano, J. P., Wiemer-Hastings, K., Millis, K. K., Muñoz, B. D., & McNamara, D. S. (2002). Using latent semantic analysis to assess reader strategies. *Behavior Research Methods, Instruments, and Computers*, *34*, 181–188.

McNamara, D. S. (2001). Reading both high coherence and low coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, *55*, 51–62.

McNamara, D. S. (2004a). Aprendiendo a través de los textos: Efectos de la estructura del texto y las estrategias de los lectores. *Signos*, *37*, 123–159.

McNamara, D. S. (2004b). SERT: Self-Explanation Reading Training. *Discourse Processes*, *38*, 1–30.

McNamara, D. S. (2006). *SERT improves low-knowledge students' science course performance*. Manuscript submitted for publication.

McNamara, D. S., Best, R., & Castellano, C. (2004). *Learning from text: Facilitating and enhancing comprehension*. Retrieved February 26, 2004, from http://www.speechpathology.com

McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (in press). Using LSA and word-based measures to assess self-explanations in iSTART. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of LSA*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McNamara, D. S., & Healy, A. F. (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory and Language*, *43*, 652–679.

McNamara, D. S., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, and Computers*, *36*, 222–233.

McNamara, D. S., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, *34*, 147–171.

McNamara, D. S., & Scott, J. L. (1999). Training reading strategies. In *Proceedings of the 21st Annual Meeting of the Cognitive Science Society* (pp. 387–392). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McNamara, D. S., & Shapiro, A. M. (2005). Multimedia and hypermedia solutions for promoting metacognitive engagement, coherence, and learning. *Journal of Educational Computing Research, 33,* 1–29.

Meyer, B. J. F., Middlemiss, W., Theodorou, E. S., Brezinski, K. L., McDougall, J., & Bartlett, B. J. (2002). Older adults tutoring fifth-grade children in the structure strategy via the Internet. *Journal of Educational Psychology*, *94*, 486–519.

Millis, K. K., Kim, H. J., Todaro, S., Magliano, J., Wiemer-Hastings, K., & McNamara, D. (2004). Identifying reading strategies using latent semantic analysis: Comparing semantic benchmarks. *Behavior Research Methods, Instruments, and Computers*, *36*, 213–221.

Millis, K., Magliano, J., Wiemer-Hastings, K., Todaro, S., & McNamara, D. S. (2005). Assessing comprehension with latent semantic analysis. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of LSA*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

National Assessment for Educational Progress. (2003). *The nation's report card: Reading highlights* (NCES Report No. 2004452). Jessup, MD.

Oakhill, J., & Yuill, N. (1996). Higher order factors in comprehension disability: Processes and remediation. In C. Cornaldi & J. Oakhill (Eds.), *Reading comprehension* *difficulties: Processes and intervention* (pp. 69–92). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

O'Reilly, T., Best, R., & McNamara, D. S. (2004). Self-Explanation Reading Training: Effects for low-knowledge readers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1053–1058). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

O'Reilly, T., & McNamara. D. S. (2006). Reversing the Reverse Cohesion Effect: Good texts cam be better for strategic, high-knowledge readers. *Discourse Processes, 43,* 121–152.

O'Reilly, T. P., Sinclair, G. P., & McNamara, D. S. (2004a). iSTART: A Web-based reading strategy intervention that improves students' science comprehension. In D. G. Kinshuk, Sampson, & P. Isaías (Eds.), *Proceedings of the IADIS International Conference on Cognition and Exploratory Learning in Digital Age* (pp. 173–180). Lisbon, Portugal: IADIS Press.

O'Reilly, T. P., Sinclair, G. P., & McNamara, D. S. (2004b). Reading strategy training: Automated versus live. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1059–1064). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Ozuru, Y., Best, R., & McNamara, D. S. (2004). Contribution of reading skill to learning from expository texts. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1071–1076). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction*, *2*, 117–175.

Paris, S., Cross, D., & Lipson, M. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology*, *76*, 1239–1252.

Perfetti, C. (1985). *Reading ability*. New York: Oxford University Press.

Pressley, M., & Ghatala, E. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist*, *25*, 19–33.

Pressley, M., Wood, W. W., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist*, *27*, 91–100.

Rothkopf, E. Z. (1988). Perspectives on study skills training in a realistic instructional economy. In C. E. Weinstein, E. T. Goetz, & P. A. Alexander (Eds.), *Learning and study strategies: Issues in assessment, instruction, and evaluation* (pp. 275–286). San Diego, CA: Academic.

Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K., Fletcher, J., Brady, S., et al. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. *Scientific Studies of Reading*, *3*, 69–94.

Shapiro, A. M. (2004). How including prior knowledge as a subject variable can change outcomes of learning research. *American Educational Research Journal*, *41*, 159–189.

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.

Stothard, S. E., & Hulme, C. (1996). A comparison of reading comprehension and decoding difficulties in children. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 93–112). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Thompson, R. A., & Zamboanga, B. L. (2003). Prior knowledge and its relevance to student achievement in Introduction to Psychology. *Teaching of Psychology*, *30*, 96–101.

Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: MIT Press.

17

Reading As Thinking: Integrating Strategy Instruction in a Universally Designed Digital Literacy Environment

Bridget Dalton and C. Patrick Proctor
Center for Applied Special Technology, Wakefield, Massachusetts

In this chapter, we present an overview of our work developing universal literacy environments (ULEs). Reflecting universal design for learning principles (Rose & Meyer, 2002), a ULE is a digital reading environment that provides the learner with a variety of embedded features that are designed to support individual learning needs while being sensitive to the interactive nature of the reading process. Focused on comprehension building, ULEs allow print-challenged students to access the same texts as their typically achieving peers through text read-aloud software. Students who struggle with making meaning are supported in an apprentice model of reading strategy instruction in which scaffolds decrease as students' understanding and self-regulation improve. It is our belief that we learn the most from engaging students in the margins of the achievement distribution, and we therefore have pursued projects that target struggling readers, including students who are learning English as a second language, students who are deaf and hard of hearing, and children with significant cognitive disabilities.

Maria: I don't really like to read.
Jay: I can't read fast enough to keep up.
Moesha: There are too many words I don't know.
Ricky: I don't have a clue what this means.
Paola: If I could read this in Spanish, I would understand it.

Children struggle to understand text for a variety of reasons, including lack of engagement; weak decoding and fluency skills; inadequate vocabulary and background knowledge; and ineffective strategies for setting a purpose for reading, monitoring one's understanding, and resolving problems (Lipson & Wixson, 1997; Paris, Wasik, & Turner, 1991). It is rare to have just one area of concern, and if the struggle goes on too long, motivation and engagement inevitably enter into the equation, depressing performance even further (Guthrie & Wigfield, 2000). The two most prevalent reasons for these challenges are (a) lack of access to good instruction and (b) learner differences that often interfere with learning from good instruction.

For the last several years, we have been working with colleagues at the Center for Applied Special Technology to develop and research scaffolded digital literacy environments focused on the promotion of struggling readers' comprehension, engagement, and efficacy. Using high-quality novels, chapter books, folk tales, informational texts, and picture books, we have created a series of digital multimedia hypertexts, or *universal literacy environments* (ULEs), that are embedded with learning supports and provide opportunities for interactive student responses. Like other authors featured in this volume (McNamara, Levinstein, & Boonthum, 2004; chaps. 14 and 16, this volume), we are interested in how digital learning environments can improve students' comprehension and self-regulation through the use of pedagogical agents and guided practice. We situate our work with reading strategies within a larger engagement perspective, as do Guthrie, Taboada and Coddington (chap. 10, this volume; Guthrie, Wigfield, & Perencevich, 2004), and we pay special attention to the needs and interests of struggling readers (chap. 8, this volume; Williams, Hall, & Lauer, 2004).

We have studied the use of ULEs with struggling and typically achieving readers in elementary and middle school classrooms, with the major focus on students in Grades 4 through 8. Because we are interested in reaching students who are functioning at the margins of school success, we have applied a universal design for learning instructional design framework (Rose & Dalton, 2002; Rose & Meyer, 2002) to the development of ULEs for students with diverse learning needs, including struggling readers and students with learning disabilities (Dalton, Pisha, Eagleton, Coyne, & Deysher, 2002), students with significant cognitive disabilities (Coyne & Dalton, 2005; O'Neill & Dalton, 2002), students who are deaf and hard of hearing (Dalton, Shlepper, Kennedy, Lutz, & Strangman, 2005), and students who are learning English as a second language (Proctor, Dalton, & Grisham, in press). In this chapter, we describe our instructional design framework, share results from our ongoing research in classrooms, and propose a model of reading comprehension that is sensitive to both traditional reading comprehension processes and the affordances possible in digital environments.

THEORETICAL AND EMPIRICAL FOUNDATIONS

Reading Comprehension

The RAND Reading Study Group (Snow, 2002) defines *reading comprehension* as a "process of simultaneously extracting and constructing meaning through interaction and involvement with written language" (p. 11). Meaning is constructed as a transaction among reader, text, and activity, situated within a larger sociocultural context. The RAND report acknowledges the important role of affect and the changing nature of the text in digital multimedia environments, aspects that we believe are key to advancing our understanding of comprehension and effective instructional practice. In a digital context, the relationship among reader, text, and activity can be changed in ways that extend the capacity of the reader and transform the text to take on teaching and learning roles (McKenna, 1998; Strangman & Dalton, 2005). For example, a seventh-grade student reading on a fourth-grade level because of decoding difficulties may productively engage with texts at his grade level if the texts are in a digital format that includes a read-aloud tool to compensate for weak word recognition skills. Digital texts may also be transformed to offer teaching supports, such as pedagogical agents who model reading processes. The text may also learn about the reader as it collects and analyzes performance data and makes adjustments in the learning environment accordingly. In effect, the reader–text–activity relationship is dynamic in a digital context.

Our work draws on the extensive body of research supporting strategy instruction (National Reading Panel, 2000) and, specifically, Reciprocal Teaching (RT; Palincsar & Brown, 1984), as well as research on hypertext comprehension and digitally supported literacy environments (for a review, see Dalton & Strangman, 2006), multimedia learning (Mayer, 2001), and engagement (chap. 10, this volume; Lepper, 1985). Our ULE prototypes include multiple strategies and supports, with a focus on strategic comprehension and engagement. They fit best within a cognitive apprenticeship model of learning (Cognition and Technology Group, 1993) and an interactive view of reading that ensures bottom-up processes through text-reading technology, while privileging top-down processes so crucial to comprehension building (e.g., Graesser, Singer, & Trabasso, 1994; Rumelhart, 1994). We are also guided by a universal design for learning framework, which we describe in the next section.

A UNIVERSAL DESIGN FOR LEARNING INSTRUCTIONAL FRAMEWORK

The concept of *universal design* originated in the field of architecture during the 1970s in response to a federal U.S. mandate requiring that physical access be provided to individual with disabilities. Ron Mace (1998), an architect with physical disabilities, argued against the prevailing practice of retrofitting buildings and physical spaces to accommodate individuals with disabilities, suggesting instead that designers consider the needs of the broadest range of users from the beginning of the planning process, with the goal of creating something that would benefit all users. Sidewalk curb cuts and TV captioning are two examples of innovations designed for individuals with physical and sensory disabilities that are now commonplace affordances in daily life.

Universal design for learning (UDL) applies this concept to the conceptualization and implementation of curricula and instruction (Rose & Dalton, 2002; Rose & Meyer, 2002). We believe that the most effective learning environments (i.e., texts, curriculum, assessments, instructional methods, etc.) are those that consider the needs and interests of the broadest spectrum of learners from the outset. This philosophy of learning is at odds with more traditional learning approaches that advocate adapting curriculum, or developing supplemental materials, for students who have special needs. Drawing on recent advances in the neurosciences and our understanding of how the brain learns, as well as the flexibility of new digital technologies and media, UDL uses three basic design principles: (a) to support diverse recognition networks, provide multiple means of representation; (b) to support diverse strategic networks, provide multiple means of strategic learning and expression within an apprenticeship environment; and (c) to support diverse affective networks, provide multiple means of engagement (Rose & Meyer, 2002). Although it is possible to apply the tenets of UDL without technology, digital multimedia are at once flexible, scalable, and standardized in such a way as to allow for the most fruitful implementations of classroom-based UDL practices.

Guided by UDL design principles, we have developed and researched a prototype ULE that applies Palincsar and Brown's (1984) RT approach to reading strategy instruction (for a review, see National Reading Panel, 2000). RT is a well-validated approach to improving students' comprehension and self-monitoring skills through an apprenticeship model of learning. The teacher and students engage in an instructional dialogue about the text, coconstructing their understanding of the text as they apply several strategies: predicting, questioning, summarizing, and clarifying. At first, the teacher plays a lead role, demonstrating, modeling, and providing feedback as the group reads a shared text. Students then take turns leading the discussion, with the teacher gradually releasing control to students as their competence increases. The goal of RT is not to teach strategies per se but rather to apply strategies in the service of developing deep understanding. Core to RT is the notion of *scaffolding*, whereby supports are dynamically adjusted to meet the needs of the learner in relation to the demands of the task (Vygotsky, 1978; Wood, Bruner, & Ross, 1976). A screen shot from a ULE folk tale is shown in Figure 17.1.

In the following section, we describe the various supports from a UDL perspective. The primary goal of the ULE is to develop engaged, active, and strategic readers who are able to make sense of complex language in a variety of educational content domains. A second important goal is to support students' access to and progress in the general education curriculum, as mandated by the Individuals with Disabilities Education Act (1997). To accomplish this latter goal, we have selected texts that are age and grade level appropriate and have embedded learning supports within these texts. At first glance, this appears contradictory to the prevailing recommendation that struggling readers should be taught with texts that are at their independent and instructional reading levels (Allington, 2001). For adolescents who are reading substantially below grade level, this commonly entails the use of high-interest, easy reading materials (which are often anything but interesting). Of course, we agree that students need to read texts at the appropriate level of text difficulty to develop the reading skills required for fluent reading and comprehension. At the same time, students must have access to grade-level texts that provide the foundation for academic learning and future success as knowledgeable, literate citizens.

The flexibility of digital text makes it possible to redefine the concept of readability by manipulating the access supports so that students can focus on making sense of the text, rather than decoding the words (Edyburn, 2002; McKenna, Reinking, Labbo, & Kieffer, 1999). A second benefit of this approach is that it directly addresses the issue of engagement. Students want to read the same books, magazines, and Web sites that their classmates are reading. Adolescents in particular find it difficult to be visibly identified as poor readers, reading different books and participating in different learning

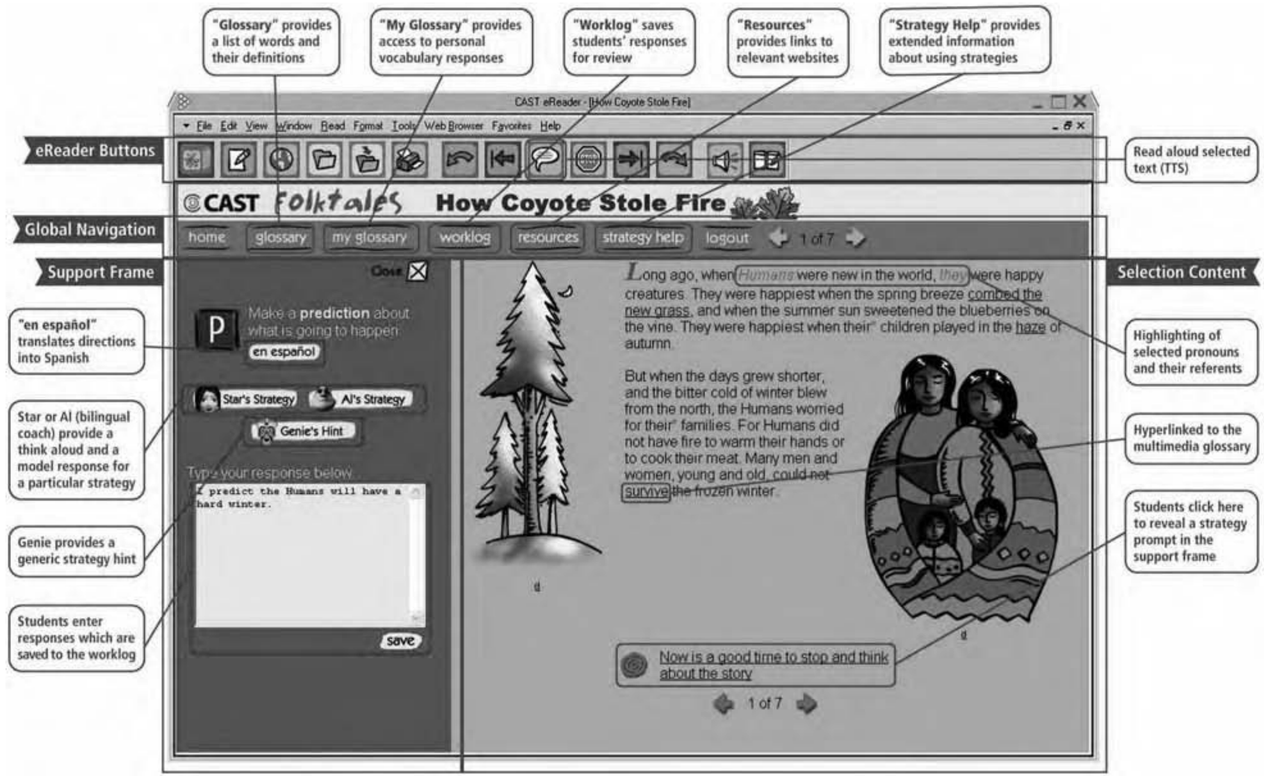


Figure 17.1. Sample screen shot of a universal learning environment folk tale with embedded supports. Reprinted with permission of the Center for Applied Special Technology, Inc., Copyright 2004.

experiences than their peers. Fink's (1995–1996) research on highly successful adults with dyslexia reveals the power of interest in driving learners to struggle with texts that were beyond their reading level to learn content that was important to them. Indeed, we hypothesize that by using age-appropriate text we are reducing some of the negative affect associated with struggling reader status, with the hope that students will more willingly persist with the challenging task of accelerating their reading growth and becoming engaged readers.

It is important to situate the use of ULEs within the larger instructional context of the classroom. Although we may work with small groups of children during the formative stages of development, field testing is carried out by teachers. Although it varies by project, teachers generally participate in a 2-day training institute, followed by occasional coaching sessions during the intervention period. Students also participate in one or two training sessions to learn about the various features and navigation system. Perhaps not surprisingly, even young children with significant cognitive disabilities quickly learn how to use the basic system.

Providing Multiple Means of Representation

It is not possible to read with understanding if you cannot recognize the words accurately and efficiently. Therefore, to support students with decoding and/or fluency issues, the ULE provides a text-to-speech (TTS) tool that allows students to click on a word, phrase, or passage and have it read aloud. The text is read with synthetic voice, accompanied by synchronized highlighting. Students can select their own voice and narration rate and are asked to read along in the text as they listen to the TTS tool so that they are both seeing and hearing the words simultaneously.

There is evidence that TTS benefits students especially when used over an extended period of time, although some research challenges this approach (for a review, see Strangman & Dalton, 2005). In our work, TTS is essential as an access tool for struggling readers who would not be able to read the text otherwise, and thus, at a minimum, it plays an important assistive technology role. In several studies, we have found that students reading ULEs improve their performance on print-based measures, suggesting that the TTS is not impeding skill development (Coyne & Dalton, 2005; Dalton et al., 2002, 2005). Although TTS is an economical means of providing read aloud functionality, there is no doubt that even the most current digital voices do not provide a good model of oral reading expression. In our work with young children, we have complemented TTS at the word level with human voice recordings at the sentence and passage level (Coyne & Dalton, 2005). From

Figure 17.2. American Sign Language video with captioning models a think-along for students who are deaf and hard of hearing (Dalton et al, 2005). Reprinted with permission of the Center for Applied Special Technology, Inc. Copyright 2005.

a universal-design perspective, it would be optimal to offer TTS and human voice so that users could select the format that best served their purpose (e.g., visually impaired individuals often prefer TTS to human voice because it is possible to listen to TTS at high speed without losing intelligibility; Jackson, May 10, 2006, personal communication).

Although a read-aloud tool is the primary means of providing access to the words, it does not meet the needs of individuals who are deaf and hard of hearing. In a recently completed project with the Laurent E. Clerc Center at Gallaudet University, Dalton and colleagues (2005) embedded American Sign Language (ASL) video and Signing Avatar clips (VCom3D) in digital texts for middle school students. Students could click on a word or passage and view it signed in ASL (see Figure 17.2). In this case, the "reading" of the word is necessarily connected to its meaning, because ASL is a fully developed language system in which the sign and meaning are interrelated (finger spelling is an exception). Just as TTS is no substitution for a human voice, the expressivity of a signing avatar is fairly restricted and thus was used for more routine messages to the student.

There are three other ways that ULEs offer multiple means of representation: multimedia vocabulary hyperlinks, background knowledge hyperlinks, and first-language translations of directions and supports for English-language learners. For example, a fifth-grade student reading a folk tale about "How Coyote Stole Fire" encounters menacing Fire Beings with talons for hands. Not sure what talons are, but curious, the student clicks on the word to obtain a definition; example sentence; and two graphics illustrating talons, one a photograph of an eagle in the wilderness and the other a diagram of a bird of prey that more clearly shows the talons. A Spanish-language translation of talons (garras) also appears, on which the student may click to hear its pronunciation in English or Spanish. Again, the goal is to provide multiple representations of "talon"—in text, with audio, with graphics, in languages other than English, and in different contexts.

Also accessible are background knowledge links, which are varied, depending on the ULE content, and are designed not only to build background knowledge that would be helpful in understanding the text but also to provide opportunities for extension and enrichment. For example, the student just described who was reading a Native American tale about the trickster Coyote could access links to Web sites to learn more about coyotes and to read online folk tales. Adolescents reading a novel set in Korea and Japan during the closing days of World War II had the option of clicking on a series of maps that situated the events in the story in their geographical location, and others reading about Martin Luther King Jr. and the March on Washington could access a biographical timeline and links to Web sites about the civil rights movement. The greatest push to expand the role of background knowledge support came through our work with young children with severe cognitive disabilities (Coyne & Dalton, 2005). We included "real life" video segments to help students with missing background knowledge and to promote text-to-life connections. For example, we embedded a video of children playing hide-and-seek in a ULE picture book about children playing this game. For these students, hide-and-seek was an unfamiliar experience, and so viewing the video helped them to not only understand the story but also led to their playing hide-and-seek in the classroom; writing their own hide-and-seek book; and in one class, creating their own hide-and-seek video. Although the children could access it at any point during their reading, the video appeared to serve an anchoring function similar to that of the introductory video used in the technology-based reading program developed by Hasselbring, Goin, and Wissik (1989).

Finally, a third type of representational support is the provision of language translations for English-language learners, specifically, students who are native Spanish speakers (Proctor et al., 2007). All directions and instructional supports are provided in Spanish, both in written and TTS read-aloud formats. One of the pedagogical agents is bilingual, and students may toggle back and forth between hearing the agent speak in Spanish or English. The bilingual coach also supports students in applying first-language knowledge to English through cognate alerts. Thus far, we have not translated the core text into Spanish, given our focus on improving English reading achievement; however, this is an avenue of needed exploration that would surely be helpful to many newcomer students in American classrooms.

Providing Multiple Means of Strategic Learning and Expression

Consistent across the various ULE research projects has been a focus on developing students' comprehension and strategic reading through embedded strategy instruction. As described earlier, we have adapted RT (Palincsar & Brown, 1984) to a digital context. In thinking about how to extend this approach to a digital literacy environment, we first made the decision to contextualize the ULE within the larger classroom context, so that teachers and students would be applying strategy instruction with print texts as well as the ULEs. Although it would be useful to develop a stand-alone program, class discussion is a key aspect of the learning experience that we wanted to maintain because we thought it would deepen interest in reading the texts and promote transfer. We are also interested in learning more about the role of the ULE as a teaching and learning tool that facilitates teachers' skill in reading comprehension strategies and enables them to differentiate instruction more effectively.

To the four RT strategies of predict, question, clarify, and summarize, we added visualization as a fifth strategy (Pressley, 2000) and a feeling response option to encourage students to make a personal connection to the text (Rosenblatt, 1978). As students individually make their way through the text, they are periodically prompted to stop and apply a strategy. They enter their response in writing or audio-recording and save it to an electronic work log that can be viewed at any time by the student and teacher. The scaffolding of the text centers on students' strategy use. For example, the folk tales ULE offers five levels of support that move students from high support to low support to independent application of the strategies, ending with an open-ended response option that can be used for any purpose (e.g., making a journal entry). The support is available "just in time" at the point of student use. We built the scaffolding system to manipulate the representation of the strategy task, students' response option, and the availability of pedagogical agents. Furthermore, we varied the level of scaffolding so that strategies that are more difficult, such as summarization, offer more support than strategies that are easier, such as prediction. This ensures that from the outset all students, including the weakest readers, are being asked to respond to open-ended prompts.

To illustrate the scaffolding framework, consider the cases of summarization and prediction. At Level 1, students are asked to select the best summary from among three responses presented in multiple-choice format. They may click on a strategy coach who offers a think-aloud or a hint coach. If students select an incorrect choice, they receive corrective feedback and are asked to try again. At Level 2, students are presented six points and asked to select the four most important points to include in the summary. Again, they may access a strategy coach or a hint coach, receive corrective feedback and a prompt to try again. At Level 3, students are asked to generate and type or audio-record their own summary into a text box and complete a self-check rubric, making revisions as needed. Because the response is open ended, the strategy coach offers a think-aloud and model response. Furthermore, key information is highlighted in the passage to help students develop their summary. At Level 4, students are prompted to choose their own strategy and type or record their response. The hint coach is present, but the strategy coach is no longer available. Finally, at Level 5, students are presented an empty text box that can be used for note-taking or other response purposes.

In contrast to summary, a lighter scaffold supports students' development of prediction. Even at Level 1, students are asked to generate their own prediction in an open-ended format. However, the coaching support varies. At Level 1, there are two strategy coaches who each offer a think-aloud and model to demonstrate that there is no single correct response, as well as the hint coach. At Level 2, one of the strategy coaches drops out, and at Level 3, the strategy coach drops a model prediction and provides only a think-aloud. Levels 4 and 5 operate just as they do with summary, with students choosing their own strategy or other response option. Examples of coaching scripts and student responses are presented in Tables 17.1 and 17.2.

In developing this scaffolding framework, trade-offs were made. We limited our use of closed-ended response options so that students were being asked to stop and think much like they would if they were reading on their own or participating in a discussion. This meant that corrective feedback was limited within the ULE and that it would be even more important to connect students' reading within the ULE to opportunities for class discussion and reflection, as well as teachers' reviews of students' work logs. Although this could be viewed as a positive feature of this approach, it is also more challenging for teachers to implement.

Although the strategy levels represented scaffolds, the use of particular support features within levels was left up to the individual student. In other words, students could self-scaffold by choosing to use the coaches, access the multimedia glossary, and have the text read aloud through a TTS tool. We anticipated that making choices about when and where to use supports would

TABLE 17. 1
Examples of Strategy Coach Thinkalouds, Models, and Hints

Question Strategy Prompt: Ask a question about something that is important to know and remember. Be sure to include the answer to your question.

Strategy coach 1	Strategy coach 2	Hint coach
Text–Specific Coaching		
Think–Aloud: Maui's brothers are still angry with him. I asked a question about that because it seemed like a major problem. *Model*: Question: What did Maui's brothers do when they saw him? Answer: When Maui's brothers saw him they paddled quickly to get away.	*Think–Aloud*: This is the first time Maui uses his magical powers so I asked a question about that. *Model*: *Question*: How did Maui catch up with his brothers' canoe? Answer: Maui used his magic to turn himself into a shark.	Try asking a question about one of the characters.
Generic Coaching		
Think–Aloud: I skimmed the text to find something really important about one of the characters. I used this information to make my question		Find a place in the passage that describes something important. Ask a question about that.

TABLE 17.2
Examples of 6th grade struggling readers' strategy responses to the novel, *So Far From the Bamboo Grove*

Strategy	*Response*
Predict	I predict Hideyo will join the war no matter what his mother said. He wants to fight for his country like his father.
Question	Where does the story take place?
Clarify	A confusing part of the passage is why Yoko can't visit the corpral. A confusing word is communist.
Summarize	Yoko and Ko and her mother were on a train. It was bombed by planes and lots of people were hurt.
Visualize	I visualize a small house in North Korea and bamboo trees everywhere surrounding the house. But a big cop comes up to the door with his shining gold badge on his coat. Then Yoko's mother comes to the door.
Feeling	I feel sorry that Yoko dosent know that Hideyo is still alive.
Self-reflect	I'm understanding things better now. Using the strategies, you can reflect on what you've read, it does help a lot.

contribute to students' self-regulation and investment in the learning. Across projects, observational data suggest that students like to make choices; however, it is also clear that some students make better choices than others and that some types of supports are viewed to be more helpful. For example, struggling readers consistently access the TTS tool to have the text read aloud. The use of coaches is more variable, with some students consulting the coaches frequently and others doing so rarely, even when it would be beneficial for them to do so. Of these three supports, the multimedia glossary hyperlinks tended to be accessed least frequently. Students rated the TTS tool and coaches as "very helpful," with glossary links as "helpful" (Dalton, unpublished data). From a practical perspective, students' choices appear to be efficient—they need the TTS tool to read text that is challenging, and the coaches provide examples of how to respond to the strategy prompts. The glossary hyperlinks are only indirectly related to their task requirements and thus may be viewed by students as tangential.

In a recent study using folk tales and partner informational texts, we added event usage tracking to the ULE so that we could more systematically investigate the relationship among support use, comprehension, and reader characteristics (e.g., struggling vs. typically achieving readers; Proctor et al., 2007). We also changed the task requirements so that vocabulary learning was a required component to see whether it would increase students' use of the glossary hyperlinks (e.g., in addition to pre- and postreading vocabulary activities, students were required to add a minimum of three words per text to their personal online glossary and explain why they had selected the word or words). Students did view more glossary items than had previously been observed, with many adding more words to their glossary than was required. Proctor et al. (2007) also noted that struggling and typically achieving readers' accessing of the strategy coaches was positively correlated with comprehension. We are continuing to investigate this issue in our current work, exploring different ways to stimulate students' strategic use of the supports. It seems likely that there will be instances when supports should be "pushed" at the learner and other times when they should be "pulled" by the learner (Dalton & Strangman, 2006). Previous research in this area has shown that students often over- or underaccess supports and that those most in need are often least strategic (Anderson-Inman, Horney, Chen, & Lewin, 1994; McKenna, 1998). In one of the few studies that investigated this issue, Reinking and Schreiner (1985) found that children benefit from having support pushed. However, a follow-up study did not find any difference between mandatory and optional presentation of support (Reinking, 1988).

Another aspect to this issue of who has control over support access, at least when applications are implemented by teachers in classrooms, is the influence of the teacher on students' views and use of the supports. For example, in a study of struggling adolescent readers' use of ULE versions of novels, we observed some teachers who used the coaches as an active teaching resource, prompting students to "click on Big Al [a strategy coach] to see what he has to say" when responding to a strategy prompt and appropriating the coaches' language in their offline teaching with print books (Dalton et al., 2002). Just as students tended to view the vocabulary hyperlinks as less helpful, so did teachers. In fact, some viewed them as an interruption to the reading process, as evidenced by one teacher aide who rather loudly admonished her student to "quit clicking on the vocabulary; just keep reading!" However, for students whose vocabulary needs were quite significant, such as middle school students who are deaf, there was frequent use of the vocabulary hyperlinks.

Providing Multiple Means of Engagement

All learning is filtered by affect (Seidel, Perencevich, & Kett, 2005). Students' perceptions about themselves as readers and learners, their ability to persist in the face of challenge, their interest in a particular topic or author, the fact that they had a fight with their best friend that morning, and so on, influence their process and outcomes for any particular reading experience. Strategic reading requires that learners set goals and a purpose for reading and monitor their progress, adapting processes and strategies in response to difficulties and task requirements. As Guthrie et al. (2004) stated, engaged readers are strategic readers. We addressed engagement in our ULE design in several ways: scaffolding supports to vary the level of challenge; providing students' choice and control over access of supports and options for response; including self-assessment so that students could reflect on their progress and set goals; connecting reading with ULEs to classroom discussion and peer interaction; and selecting authentic literature that was age-appropriate, interesting, and in service of the general education curriculum standards. In the preceding section, we described the scaffolding and choice features because they also are core to supporting strategic learning and expression. Therefore, in this section we address the roles of self-assessment, social interaction, and authentic literature.

Given our focus on helping students become more strategic and in charge of their learning, we included self-evaluation prompts in the ULE and developed a student–teacher conferencing procedure. Students review their work logs to find examples of their best work and to identify an area of improvement. They then meet with their teacher to discuss their progress, set goals, and decide together whether to move up a level. The goal is not to move the student up a level but rather to help the student become more aware of how he is thinking and responding as he reads, and to highlight strategies that are

TABLE 17.3
A Conference Between a 7th-Grade Boy With Learning Disabilities and His Resource Room Teacher About Reading Progress and Strategy Use

T—All right honey, so let's see … honestly, what do you feel about the book?
Student—I like the parts about the war.
T—You like the war parts. OK, but … is there a but?
Student—No, the book's OK.
T—Good, that makes me happy. OK, now what do you think is your best work?
Student—I don't know, it's all bad. Visualization?
T—OK, let's look at some [in the work log].
Student—Most of my questions don't even make sense.
T—Is that what you want to work on?
Student–Yeah.
T—OK, Rick, the part most people get hung up on is explaining why the question is important, you have to take the time to type in why you picked that answer. It's two parts—the right question and why it's important. Call me over if you need some help! OK, so we'll work on that–[writes on conference sheet] by calling the teacher over to me and discussing why the question is important. So then if one person thinks this part is important. We can discuss it, all right? OK … clarification is good!
Student—I think visualization is so easy, that's probably why I'm so good at it.
T—You know what, I gotta tell you though, Rick, your predictions are really good, too. [reads one aloud from the work log]. See, this is a very good prediction. [reads some more examples aloud]. I'm gonna write, "awesome predictions, great visualizations, and wants to work on questioning." Good for you, picking out one you want to work on and talking about it. OK, you're on Red [level], do you want to move up? You don't have to.
Student—I'll stay at Red.
T—OK, good job, Rick.

particularly useful. A conference between a seventh-grade boy with learning disabilities and his teacher is illustrated in Table 17.3.

We view students' self-assessment as core to the learning process and the long-term goal of developing readers who not only feel in charge of their own learning but also feel they have the skills needed to accomplish their goals. That said, teachers found it challenging to find the time to review students' work logs, which could contain several pages of response, and meet with them to discuss their progress. Students also found it challenging because they were unaccustomed to reflecting on their learning and setting goals. We are currently exploring ways in which we can build teacher evaluation and response options into the student work log environment so that teachers can use this qualitative data more efficiently. We also are embedding additional objective measures, such as comprehension quick checks, that can give them a sense of how students are comprehending the text, and we are expanding the ways in which we will provide students information about their performance to support their self-evaluation and help them make better decisions about when to access the various learning features.

Across projects, we have used authentic, age-appropriate texts that we thought would appeal to students. In some cases, we selected award-winning books; in other cases, we created texts on topics that are of general interest to students of the targeted age and field-tested them to obtain student feedback. Classroom observations and student and teacher interviews indicate that students find the texts interesting and that this is part of what motivates them to keep reading. Although we have some dramatic examples of students who were transformed into readers once they were able to experience success reading in a supported environment, the fairly limited numbers of ULE texts does not allow us to explore the role of text choice, or the effect of text quantity, both factors that we know are important to students' reading development and engagement.

DISCUSSION AND CONCLUSION

A Framework for Reading Comprehension in the Digital Age

The goal of the ULE is to ensure facile interaction between the UDL-defined recognition, strategic, and affective networks in the service of improved reading comprehension outcomes. To this end, our work in literacy development centers around the hypothesized association between the development of oral and written vocabulary (Nagy, 1988), reading comprehension strategy use (Palincsar & Brown, 1984), and reading comprehension. In an ideal situation, our ULEs facilitate this interaction through the following three steps.

1. *Easing and promoting recognition networks for decoding skills and oral language development*. Decoding skills difficulty is managed through TTS supports, which allow the student to focus on the comprehension demands at hand. Indeed, observations of struggling readers using our ULEs support the notion that children who have trouble with decoding English text access TTS supports with a high degree of frequency. Recognition demands are enhanced for oral language and reading vocabulary development as well. By providing multiple representations of words through example sentences, images, definitions, and Spanish-language translations, students have multiple means of tapping a core understanding of word meanings.
2. *Developing strategic networks focused on comprehension and word learning strategies*. In our ULEs, students consistently stop and think about what they are reading and respond to a targeted think-aloud such as "Make a prediction about what will happen next in the story." Students respond through a written or oral medium, and their work is

then stored for future viewing and/or sharing with classmates. Before, during, and after reading, students may also engage in targeted word learning activities designed to increase semantic and morphological understandings between and among words, such as completing online semantic word webs and building a personalized multimedia glossary. Activities such as these serve to promote deliberate choice-making about individual learning needs based on interest and perceived strengths and needs.

3. *Providing engaging digital learning environments*. ULEs are designed for interactivity and student engagement; however, we do not pretend to simulate a video game environment. Although some theorizing is being done in the arena of human learning and gaming (see, e.g., Gee, 2003), we do not expect to achieve such a hypnotic level of student engagement. Indeed, a ULE is designed to provide motivation and engagement within the system but, unlike gaming environments, the learning that occurs in a ULE is meant to be brought back and applied broadly in the classroom across the content areas.

Furthermore, recognizing that the "future is in the margins" (Meyer & Rose, 2005, p. 13), and believing that society, as well as individuals, derive great benefit from inclusiveness, we have applied this basic universal design literacy framework with young children with significant cognitive disabilities (Coyne & Dalton, 2005), students who are deaf and hard of hearing (Dalton et al., 2005), Spanish–English bilingual students (Proctor et al., in press), and other struggling readers (Dalton et al., 2002). The results have been promising, and we are expanding this work in several ways. First, we are integrating digital scaffolded reading environments with scaffolded multimedia composition environments so that students are learning with, through, and about language and media in receptive and expressive forms, situated in a larger communication space. Second, we have developed a Web browser tool to support students' strategic reading and viewing on the Internet, where text is defined by hypertext structure, multimedia, interactivity, and social networking. Although the Internet poses new demands on the reader, such as a heightened need for critical analysis and speed of processing (Leu, Kinzer, Coiro & Cammack, 2004), it also offers untold affordances that could potentially level the playing field for diverse learners and support individuals in developing and sharing their knowledge, skills, and talents (Dalton & Strangman, 2006; Palincsar & Dalton, 2005). Finally, we are broadening our view of what constitutes a scaffolded literacy environment to include more supports for teachers, so that they too are able to get just-in-time support to help them better understand, assess, and teach children to become strategic, engaged readers and learners.

ACKNOWLEDGMENTS

This research was supported in part by grants to CAST. Inc. from the U.S. Department of Education Institute of Education Sciences and the U.S. Department of Education Office of Special Education Programs.

REFERENCES

Allington, R. (2001). *What really matters for struggling readers: Designing research-based programs*. New York: Longman.

Anderson-Inman, L., Horney, M., Chen, D., & Lewin, L. (1994). Hypertext literacy: Observations from the ElectroText project. *Language Arts*, *71*, 279–287.

Cognition and Technology Group (1993). Examining the cognitive challenges and pedagogical opportunities of integrated media systems: Toward a research agenda. *Journal of Special Education Technology*, *12*, 118–124.

Coyne, P., & Dalton, B. (2005, December). *Literacy by design: A universally designed reading environment for students with severe cognitive disabilities*. Paper presented at the 55th National Reading Conference, Miami, FL.

Dalton, B. (2004). *Strategic Learning Editions Project and Engaging the Text Project.* Unpublished raw data.

Dalton, B., Pisha, B., Eagleton, M., Coyne, P., Deysher, S. (2002). *Engaging the text: Reciprocal teaching and questioning strategies in a scaffolded learning environment* (Final report to the U.S. Department of Education, Office of Special Education Programs). Peabody, MA: Center for Applied Special Technology.

Dalton, B., Shlepper, D., Kennedy, M., Lutz, L., & Strangman, N. (2005). *Chapter by chapter: Thinking reader for students who are deaf and hard of hearing* (Final report to Gallaudet University and the Laurent E. Clerc National Center on Deaf Education). Wakefield, MA: Center for Applied Special Technology.

Dalton, B., & Strangman, N. (2006). Improving struggling readers' comprehension through scaffolded hypertexts and other computer-based literacy programs. In D. Reinking, M. C. McKenna, L. D. Labbo, & R. D. Keiffer (Eds.), *Handbook of literacy and technology* (2nd ed., pp. 75–92). Mahwah, NJ: Lawrence Erlbaum, Inc.

Edyburn, D. L. (2002, April–May). Cognitive rescaling strategies: Interventions that alter the cognitive accessibility of text. *Closing the Gap*, *1*, 10–11, 21.

Fink, R. P. (1995–1996). Successful dyslexics: A constructivist study of passionate interest reading. *Journal of Adolescent and Adult Literacy*, *39*, 268–280.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Macmillan.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.

Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Volume III* (pp. 403–422). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Guthrie, J. T., Wigfield, A., & Perencevich, K. C. (Eds.). (2004). *Motivating reading comprehension: Concept-oriented reading instruction*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hasselbring, T. S., Goin, L. I., & Wissik, C. (1989). Making knowledge meaningful: Applications of hypermedia. *Journal of Special Education Technology*, *10*, 61–72.

Individuals with Disabilities Education Act Amendments of 1997. Pub. L. 105-17, 20 U.S.C. § 1400 et seq.

Lepper, M. R. (1985). Microcomputers in education: Motivational and social issues. *American Psychologist*, *40*, 1–18.

Leu, D. J., Jr., Kinzer, C. K., Coiro, J. L., & Cammack, D. W. (2004). Toward a theory of new literacies: Emerging from the Internet and other information and communication technologies. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1570–1613). Newark, DE: International Reading Association.

Lipson, M. Y., & Wixson, K. K. (1997). *Assessment & instruction: An interactive approach*. New York: Longman.

Mace, R. L. (1998). Universal design in housing. *Assistive Technology*, *10*, 21–28.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge, England: Cambridge University Press.

McKenna, M. (1998). Electronic text and the transformation of beginning reading. In D. Reinking, M. C. McKenna, L. D. Labbo, & R. D. Kieffer (Eds.), *Handbook of literacy and technology: Transformations in a post-typographic world* (pp. 45–60). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McKenna, M. C., Reinking, D., Labbo, L. D., & Kieffer, R. D. (1999). The electronic transformation of literacy and its implications for the struggling reader. *Reading & Writing Quarterly*, *15*, 111–126.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, and Computers*, *36*, 222–233.

Meyer, A., & Rose, D. H. (2005). The future is in the margins: The role of technology and disability in educational reform. In D. H. Rose, A. Meyer, & C. Hitchcock (Eds.), *The universally designed classroom: Accessible curriculum and digital technologies* (pp. 13–35). Cambridge, MA: Harvard Education Press.

Nagy, W. (1988). *Teaching vocabulary to improve reading comprehension*. Newark, DE: International Reading Association.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Jessup, MD: National Institute for Literacy.

O'Neill, L. M., & Dalton, B. (2002). Thinking readers Part II: Supporting beginning reading in children with cognitive disabilities through technology. *Exceptional Parent*, *32*(6), 40–43.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition & Instruction*, *1*, 117–175.

Palincsar, A. S., & Dalton, B. (2005). Speaking literacy and learning to technology: Speaking technology to literacy and learning. In B. Maloch , J. Hoffman, D. Schallert, C. Fairbanks, & J. Worthy (Eds.), *54th yearbook of the National Reading Conference* (pp. 83–102). Oak Creek, WI: National Reading Conference.

Paris, S. G., Wasik, B. A., & Turner, J. C. (1991). The development of strategic readers. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (pp. 609–640). White Plains, NY: Longman.

Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 545–562). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Proctor, P., Dalton, B., & Grisham, D. (2007). Scaffolding English language learners and struggling readers in a digital environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research*, *39,* 71–93.

Reinking, D. (1988). Computer-mediated text and comprehension differences: The role of reading time, reader preference, and estimation of learning. *Reading Research Quarterly*, *23*, 484–498.

Reinking, D., & Schreiner, R. (1985). The effects of computer-mediated text on measures of reading comprehension and reading behavior. *Reading Research Quarterly*, *20*, 536–552.

Rose, D., & Dalton, B. (2002). Using technology to individualize reading instruction. In C. C. Block, L. B. Gambrell, & M. Pressley (Eds.), *Improving comprehension instruction: Rethinking research, theory, and classroom practice* (pp. 257–274). San Francisco: Jossey-Bass.

Rose, D., & Meyer, A. (2002). *Teaching every student in the digital age: Universal design for learning*. Alexandria, VA: Association for Supervision and Curriculum Development.

Rosenblatt, L. M. (1978). *The reader, the text, the poem: The transactional theory of the literary work*. New York: Appleton-Century.

Rumelhart, D. E. (1994). Toward an interactive model of reading. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 864–894). Newark, DE: International Reading Association.

Seidel, R. J., Perencevich, K. C., & Kett, A. L. (2005). *From principles of learning to strategies for instruction: Empirically based ingredients to guide instructional development*. New York: Springer.

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND Corporation.

Strangman, N., & Dalton, B. (2005). Technology for struggling readers: A review of the research. In D. Edyburn, K. Higgins, & R. Boone (Eds.), *The handbook of special education technology research and practice* (pp. 545–569). Whitefish Bay, WI: Knowledge by Design.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Williams, J. P., Hall, K. M., & Lauer, K. D. (2004). Teaching expository text structure to young at-risk learners: Building the basics of comprehension instruction. *Exceptionality*, *12*, 129–144.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *17*, 89–100.

18

Designing a Hypermedia Environment to Support Comprehension Instruction

Annemarie Sullivan Palincsar
University of Michigan

Rand J. Spiro
Michigan State University

Linda Kucan
University of Pittsburgh

Shirley J. Magnusson
California Polytechnic State University

Brian Collins
Michigan State University

Susanna Hapgood
University of Toledo

Aparna Ramchandran
Michigan State University

Nancy DeFrance and Adrienne Gelpi-Lomangino
University of Michigan

Although the research literature suggests a range of ways in which teachers might enhance students' ability to interpret and learn from text, comprehension instruction continues to be largely absent in classroom-based reading curricula and pedagogy. In this chapter, we describe how we designed a hypermedia system specifically for the purpose of supporting teachers to learn two approaches to comprehension instruction for which there is a robust evidentiary base: (a) Questioning the Author and (b) Reciprocal Teaching. The system is organized around a corpus of video clips that can be searched and arranged according to 10 themes germane to planning and enacting comprehension instruction. We describe how we: drew on cognitive flexibility theory to design the system as an experience acceleration support environment, used the knowledge base regarding text comprehension and teacher learning of complex information, and (c) designed the professional development context to optimize use of the system.

This volume is replete with knowledge claims about the processes of understanding and learning from text that, directly or indirectly, have implications for teaching. The focus of this chapter is on the design of a hypermedia environment intended to support the learning of teachers about a number of complex issues that attend reading comprehension instruction. Our goal in this chapter is to describe the design decisions and how they were informed by the knowledge base regarding two things: (a) text comprehension and effective comprehension instruction and (b) learning from hypermedia systems. What motivates our interest in presenting this information is an awareness that there is a scholarship behind the design of learning environments that is seldom transparent and often unacknowledged. Before describing this hypermedia environment, we characterize the problems of upper elementary students that are core to our work, and we then describe the problems of teaching reading comprehension.

STUDENTS' DIFFICULTIES WITH TEXT COMPREHENSION

Significant numbers of students who read well enough in the primary grades (i.e., who acquire adequate decoding skills and can interpret simple narrative text) demonstrate a decline in their rate of reading progress around the fourth grade (RAND, 2002). This phenomenon has been referred to as the *fourth-grade slump* (Chall, Jacobs, & Baldwin, 1990). An analysis of fourth-graders' performance on the 2003 National Assessment of

Educational Progress (NAEP) reading assessment revealed that 37% could not make simple inferences based on text information and prior knowledge (NAEP 2003). Furthermore, although NAEP mathematics scores have improved over the past 30 years, reading scores have declined (Donahue, Voekl, Campbell, & Mazzeo, 1999). Particularly troubling is the finding that the largest and most persistent gaps between grade level and achievement were for poor, urban youth.

Explanations for the decline in reading achievement that students demonstrate in the upper elementary grades have focused on the changing nature of the demands in these grades. For example, there is a change in the purposes for which children are expected to read, which Chall (1983) described as the shift from "learning to read" to "reading to learn." There is also a change in the genre of text that students experience. Whereas students in the primary grades are exposed almost exclusively to narrative text (Duke, 2000), the use of expository text increases as students enter the upper elementary grades (Hiebert & Fisher, 1990; Donovan & Smolkin, 2001; Taylor, Pearson, Peterson, & Rodriguez, 2002). This change is significant because the kinds of skills and strategies required for understanding expository text differ from those required for narrative text (Goldman, 1997; Spiro & Taylor, 1987; Wilson & Rupley, 1997). For example, McNamara, Floyd, Best, and Louwerse (2004) found that whereas reading skill is a predictor of success in understanding narrative text, prior knowledge is the better predictor of success with expository text. Furthermore, Palincsar et al. (2004) found that both typical and struggling readers used their knowledge of story structure (e.g., characters, plot, and resolution) to bring coherence to their reading of narrative text but had nothing analogous on which to draw relative to bringing coherence to their reading of exposition.

Recently, significant federal and state reading initiatives, such as Reading First, have been launched to improve the reading achievement of children in the United States. These initiatives have primarily focused on decoding instruction. Clearly, attention to such instruction is well placed; we know that sound early reading instruction can make a significant difference in the reading achievement of young children (Snow, Burns, & Griffin, 2000). However, we also know that sound early reading instruction must address not only the decoding of text but also text comprehension. A recent cluster analysis conducted by Buly and Valencia (2002) on the reading scores of fourth-graders in the state of Washington supports this claim. They analyzed the reading and language profiles of the 43% of fourth-graders in their sample who failed to meet proficiency on the statewide assessment. More than 50% of these students showed evidence of solid basic skills in word identification but struggled with comprehension and fluency.

THE CHALLENGES OF TEACHING READING COMPREHENSION

Meta-analyses reported by the National Reading Panel (2000); Swanson and Hoskyn (1998); Gersten, Williams, Fuchs, and Baker (1998), and many of the authors in this volume suggest that a good deal is known about specific strategies that enhance comprehension, effective instruction to teach those strategies, and the relationship between reading strategy instruction and engagement and learning (see also Pressley, 2000). In fact, across the meta-analyses cited above, the greatest effect sizes have been from reading strategy instruction.

Why, then, has there not been more success advancing student achievement with respect to comprehending expository text? One explanation is that, despite the research knowledge, teachers report feeling inadequately prepared to teach reading comprehension, especially with students who struggle in school (Bryant, Linan-Thompson, Ugel, Hamf & Hougen, 2001; Vaughn, Hughes, & Schumm, 1998). This lack of preparation is acknowledged in compelling terms in the report of the RAND Reading Study Group (RRSG): "The RRSG believes that teachers must be front and center in the discussion about how to improve comprehension instruction in schools today" (RSSG, 2002, p. xviii). The RRSG report goes on to emphasize the complexity of teaching reading comprehension, a point underscored by the National Reading Panel (2000):

> What we must remember is that reading comprehension instruction is extremely complex and that teaching reading comprehension is also extremely complex. The work of researchers discussed here makes this clear. They have not recommended an "instructional package" that can be prescribed for all students. They have not identified a specific set of instructional procedures that teachers can follow routinely. Indeed, they have found that reading comprehension cannot be routinized. (pp. 4–125)

The complexity of comprehension instruction should not be reduced during professional development experiences to ensure that teachers' instruction does not become simply procedural and mechanical and, consequently, ineffectual (Palincsar, Stevens, & Gavelek, 1989). In essence, research indicates that teachers need to know not only the specific practices, but also the principles that underlie those practices, so that they can appropriately adapt and modify the practices as needed in their local context. However, it is difficult to render the principles of effective comprehension instruction accessible to teachers in the relatively short time spans typically allocated for professional development.

RECIPROCAL TEACHING AND QUESTIONING THE AUTHOR

Two contemporary approaches to text comprehension instruction bring into relief the challenges identified above: Reciprocal Teaching (RT) and Questioning the Author (QtA). Because these approaches are the centerpiece of the hypermedia environment, we consider each in turn.

Reciprocal Teaching: The Method

RT (Palincsar & Brown, 1984) is an instructional procedure designed to enhance students' reading comprehension. The procedure engages teachers and students in a dialogue aimed at determining the meaning of the text. The dialogue is supported by the use of four strategies: (a) question generating, (b) summarizing, (c) clarifying, and (d) predicting. The teacher initially leads the dialogue and models the application of these strategies to bring meaning to the written word and monitor one's own thinking and learning from text. Over time, the students assume increased responsibility for leading the dialogues.

RT contrasts with traditional reading instruction, in which the strategies are presented as a set of isolated skills and are seldom practiced in the actual context in which they will be used. Isolated practice of individual strategies does not reinforce their flexible and opportunistic use. Hence, traditional reading instruction has typically provided inadequate opportunities for children to acquire knowledge about when to use particular comprehension strategies. One other advantage of RT is that it takes into consideration the influence of motivation on student learning and the kinds of attributions that students who have a history of academic difficulty typically make. RT enhances motivation because students typically enjoy interacting with their peers and collaborating with their teachers, and it increases student awareness of the kinds of factors that influence learning outcomes. Finally, as students become experienced with RT dialogues, they come to appreciate the relationship between their activity as readers and the outcomes of this activity. The principal challenges teachers confront in effectively implementing RT include supporting students to make opportunistic and flexible use of the strategies, learning a repertoire of ways to support (scaffold) the participation and learning of struggling readers, and understanding that the dialogues develop over time and that teachers' practices need to be coordinated with this development.

Reciprocal Teaching: The Effectiveness

Approximately 300 middle school students and 400 primary grade students participated in Palincsar and Brown's (1984) experimental research on RT.

Designed especially for students who were at risk for academic difficulty, or who were already identified as remedial or special education students, these participants typically scored below the 40th percentile on nationally normed measures of reading achievement. To evaluate the success of the intervention, criterion-referenced measures of text comprehension were administered. These assessments were designed to evaluate students' ability to recall information, draw inferences, identify the gist of the passage, and apply information presented in the text to a novel situation. Compared with the control condition, students in the RT condition (20 days of instruction) exhibited statistically and educationally significant gains. Furthermore, participants maintained these gains for up to 6 months to 1 year after the instruction (Palincsar & Brown, 1984, 1989).

Rosenshine and Meister (1994) conducted a meta-analysis of 16 studies of RT, involving participants from age 7 to adulthood, in which RT was compared with: traditional basal reading instruction, explicit instruction in reading comprehension, and reading and answering questions. Using standardized measures to assess comprehension, they determined the median effect size to be 0.32, significantly favoring RT. Experimenter-developed comprehension tests resulted in a median effect size of 0.88. Furthermore, the researchers found no significant difference in outcomes based on number of instructional sessions (which ranged from 6 to 25) or on the size of the instructional group (which ranged from 2 to 23). In closing, the National Research Council (Donovan, Wigdor, & Snow, 2003), recently identified RT as one of several research-based practices that should be supported for broad-scale implementation.

Questioning the Author: The Method

QtA is a procedure that, like RT, takes the form of a discussion that unfolds as students are reading a text for the first time and collaborate in determining the meaning of the text (Beck, McKeown, Hamilton, & Kucan, 1998; Beck, McKeown, Worthy, Kucan, & Sandora, 1996). Two sets of tools are featured in QtA: (a) queries and (b) discussion moves (Beck et al., 1997). The use of queries in QtA is a complement to the strategies of RT. Queries serve a broad range of purposes, including promoting interaction with the text, grappling with ideas, investigating with the text, and clarifying ideas in the text. Challenges particular to the successful enactment of QtA include listening to students and weaving their comments into meaningful discussion, crafting questions that are both supportive and open, and identifying the targeted big ideas in text.

Questioning the Author: The Effectiveness

One set of studies has investigated the use of QtA with over 120 students in the upper elementary grades. These studies compared of the nature of the discourse in reading and social studies classes before and after implementation of QtA. Beck et al. (1996) reported the following three improvements (among others) in classroom discourse: (a) question asking that focuses on constructing and extending meaning, rather than mere information retrieval; (b) teacher responsiveness to students that is designed to deepen the probing of the text, in addition to evaluating student contributions; and (c) frequent student initiation of questions. Sandora (1994) compared QtA with another discussion method in which sixth- and seventh-grade students first read and then discussed whole text selections. The results indicated that the students in the QtA condition recalled more from each selection and more successfully responded to inference questions.

The complexities of these two forms of comprehension instruction, in hand with the evidentiary base supporting them as effective means of text comprehension instruction, render QtA and RT ideal candidates for professional development in the area of comprehension instruction. In addition, these two forms of instruction work in complementary ways. Both are designed to engage teachers and students in the collaborative construction of the meanings of text; however, in RT, which was principally designed for use with struggling comprehenders, the strategies are prominent in supporting the discussion (particularly in the initial days of instruction) and, initially, the content plays a less prominent role in shaping the discussion, whereas in QtA, which is conducted as whole-class instruction, the content is always at the centerpiece of the discussion. Whereas RT is purposefully planned as a procedure in which the teacher plays a less significant role over time, QtA is not designed in this fashion (although there is evidence that students do indeed internalize the queries and play a more prominent role in shaping the conversation over time). In the next portion of this chapter, we present the argument for using a hypermedia environment as the platform for supporting professional development regarding these two instructional approaches.

USE OF HYPERMEDIA TO SUPPORT TEACHERS' LEARNING OF COMPLEX INSTRUCTION

Scholars who have investigated educational reform efforts have argued that the most effective reform efforts are those that target instructional change, as opposed to, for example, policy or school climate (Elmore, 1996; Fullan, 2000). That said, it has also been acknowledged that innovation efforts that focus on instruction are the most challenging (Meyer & Rowan, 1978; Tyack & Cuban, 1995), requiring that teachers experience opportunities to learn that (a) are situated in the curriculum in a manner consistent with the research base and that are focused on student learning of specific information (Cohen & Hill, 2000), (b) provide concrete resources that will both facilitate and extend teacher learning (Borko, Davinroy, Bliem, & Cumbo, 2000); and (c) provide occasions for teachers to be engaged in active learning (Garet, Porter, Desimone, Birman, & Yoon, 2001). Our work was guided by an awareness of these attributes of professional development.

A number of teacher educators and researchers have looked to video as an important means of attaining these features. Specifically, it has been argued that video provides insights into classroom complexity (Richardson & Kyle, 1999; Spiro, Coulson, Feltovich, & Anderson, 1988/2001), is a useful means of sharing exemplary practice to support enactment of specific curricula/instruction (Lampert & Ball, 1998), provides a common reference for collaborative discussion and personal reflection (Hewitt, Pedretti, Bencze, Vallaincourt, & Yoon, 2003; Lundeberg, Levin, & Harrington, 2000), provides an opportunity to examine instruction from multiple perspectives (Miller, Kelly, & Zhou, 2004; Putnam & Borko, 2000), and encourages evaluative versus descriptive noticing of teaching practice (Sherin & Han, 2004).

Video technologies are seen as important tools for creating contexts in which teachers can engage in collaborative, multilevel, and sustained investigations of teaching, with the capacity to present richer and more detailed teaching situations than narrative cases. Furthermore, the larger array of cases that can generally be accessed in a hypermedia system, as opposed to the use of linear, discrete video, is hypothesized to reduce the danger that learners will consider the case content as prescriptive in nature, as opposed to illustrative of sets of possibilities. Finally, hypermedia systems offer additional resources, such as curricular materials, interviews with teachers, and work samples from students, that can be linked with the video clips, enhancing the set of tools with which the viewer is working.

Despite the enthusiasm for video and hypermedia technologies to enhance teachers' practices, their effectiveness is, as Wang and Hartley (2003) concluded from their careful review of this literature, "more often assumed than carefully documented" (p. 105). In the case where careful research has been conducted, important cautions emerge. For example, Hughes, Packard, and Pearson (1997) investigated the use of Reading Classroom Explorer, a hypermedia system that offers an array of video clips illustrating a broad range of literacy instructional practices in a diverse array of contexts and has been studied in the context of preservice teacher education. These researchers identified the following two challenges when using video to support teacher learning: (a) their participants used the video to support what they already knew, pointing to the strong impact of viewers' prior knowledge and experience in their viewing, and (b) the viewers, although facile at reporting what they saw in each video case, did not engage in cross-case analyses, as intended by the developers.

Mindful of the opportunities and challenges associated with the use of hypermedia environments as a site for professional development, specific to instructional practice, we undertook the design of a hypermedia environment, which we named *Experience Acceleration Support Environment: Comprehension* (EASE-C). As we describe the activities with which teachers used EASE-C, the reader will learn why this hypermedia system is an appropriate vehicle for communicating the complexities of comprehension instruction and how it was designed to support teachers' knowledge-building regarding comprehension instruction. However, first we describe this EASE system in general terms, and then we proceed to describe how this system was assembled, drawing on both theory and research. We then describe the contexts in which we implemented the use and study of EASE-C. EASE-C can be accessed at http://edr1.educ.msu.edu/CompStrat/login.asp (username: Demo, password: Demo), and the reader may find it useful to refer to this site while reading the following description.

EXPERIENCE ACCELERATION SUPPORT ENVIRONMENT FOR COMPREHENSION

EASE-C is one kind of EASE system. An EASE is a hypermedia learning environment based on cognitive flexibility theory (CFT) and CFT's approach to the design of hypermedia learning environments (Spiro, Collins, Thota, & Feltovich, 2003; Spiro et al., 1988/2007; Spiro, Feltovich, Jacobson, & Coulson, 1991; Spiro & Jehng, 1990). Within a structured hypermedia system, EASE-C features digital video cases illustrating key facets of teacher–student interactions during comprehension instruction. EASE-C incorporates an array of media, including video-clips, transcripts of classroom dialogue, teacher interviews, and researcher "color commentaries." Furthermore, EASE-C incorporates advanced hypermedia learning features based on CFT, including:

- Methods of selection and arrangement of cases to promote deep learning (e.g., drawing teachers' attention to interesting connections among cases that are not transparent, or setting up case contrasts that illustrate surprising difference where at first there seemed to be primarily similarity; see also Bransford & Schwartz, 1999, on contrasting cases).
- An ecological approach to interconnecting the conceptual themes of the instructional approaches that allows those components of instruction to find their natural patterns of co-occurrence and mutual interaction

within the natural unfolding of instructional events (e.g., occasions when a particular kind of ongoing assessment occurs; in the context of a particular kind of supportive teaching practice; for material that has a particular set of text features; all in the context of an individual child with a certain amount of vocabulary knowledge, prior knowledge, motivation); this approach contrasts with traditional video-based cases in which video clips would typically be nested under "chapters" for each separate component and are taught separately.

- A search capability that permits the user to queue video clip sequences that illustrate different combinations of the themes of RT and QtA as well as features of comprehension instruction situations; this enables the learner to construct increasingly more complex and sophisticated hypotheses about the interaction of key concepts that can then be tested against the evidence of the actual occurrences in the video (a kind of "combinatorial idea playpen"); for example, the role of scaffolding as a function of its interaction with child characteristics, text features, and conceptual complexity of the topic.
- Special features that call attention to oversimplified ways of thinking or habits of mind and then support the establishment of more appropriately complex ways of thinking; examples would include the addition of a spotlight to call attention to a subtlety that might be missed in a typical viewing of an instructional scene (e.g., a look of puzzlement passing over the face of a child), or a color wash that signals an instance of a particular feature of the instructional context (e.g., a shift in the instructional conversation initiated by a child, as opposed to the teacher).

In the EASE approach, cases are revisited in different contexts to bring out their complexity because of the difficulty in apprehending case complexity in single viewings from a single contextual perspective. Revisiting is not repeating. Because these revisited video scenes rapidly become highly familiar, it becomes possible to subsequently present them in greatly attenuated form in subsequent quick-cut "criss-crossings," exponentially increasing the number of contrasts that can be made and deeply processed in a very short amount of time, hence accelerating the rate of learning (Spiro et al., 2003). This is in contrast to what teachers might be able to experience, notice, and reflect on in the long-term course of real-time experiences with all that competes for their attention in classroom life. Thus, the acquisition of critical experience to advancing practice is accelerated with the use of EASE. Next, we discuss how we designed experiences with EASE-C to ensure that that these features were properly exploited, but first we describe the process by which we constructed EASE-C.

Construction of Experience Acceleration Support Environment for Comprehension

The construction of EASE-C began with the design, conduct, and videotaping of RT and QtA lessons and accompanying interviews with the teachers. Given that our target population was upper elementary students, we videotaped in classrooms of fourth- and fifth-graders. The RT instruction was conducted by Annemarie Sullivan Palincsar, and the QtA instruction was conducted by Linda Kucan. The RT instruction was conducted with a group of 10 students, all of whom struggled with some aspect of text comprehension (but few of whom struggled with decoding). QtA instruction, on the other hand, was conducted with entire classes of heterogeneous students. To capture change over time, RT instruction was conducted for 13 consecutive school days; we then returned 5 months later and videotaped another 7 days of instruction. A total of three QtA lessons in a total of four classes were videotaped over consecutive days. We chose to use an identical text across the two instructional approaches to support teachers' appreciating the similarities and differences in these approaches. We selected an informational (science) trade book that was both meaty enough to sustain the pursuit of meaning and challenging enough that it required thoughtful reading and discussion. Given the duration of the RT instruction, we used more chapters in the trade book and used a biography and an excerpt from a social studies text.

All of the instruction was observed by Shirley J. Magnusson and Susanna Hapgood, who served in the role of interviewer. There was a clear structure to the daily interviews, which were conducted immediately after instruction. The first part of the interview was an opportunity for the teacher to share her sense of the lesson; for example, overall impressions, critical events that stood out for the teacher, milestones in the lesson, and insights into particular children or the progress of the instruction. The second part focused on the trade-offs and dilemmas of which the teacher was aware in the course of instruction. A typical trade-off would be supporting individual student participation versus attending to issues of pacing for the group. A dilemma would be selecting a text that was sufficiently challenging, yet accessible. These trade-offs and dilemmas were particularly important to the construction of an EASE system in the sense that one goal was to make clear to teachers the ill-structured and complex nature of these forms of pedagogy. The third and final part of the interview focused on the teacher's planning for the next day; given the day's events, the teacher focused on what would be prominent in her thinking as she prepared for the next lesson. All told, we collected 25 hours of instruction and 10 hours of interviews.

After the videotaping, we began the process of organizing and coding video clips. At first, each lesson was viewed and characterized in broad strokes,

TABLE 18.1
Goals for Professional Development With EASE-C

- Learn how, when, and for whom to use QtA and RT in their comprehension instruction
- Distinguish between comprehending text and learning from text
- Identify the broad range of challenges students encounter comprehending and learning from text
- Know how to appropriately adapt instruction to the demands of the text and the profile of the reader
- Learn how to create productive instructional contexts

using narrative descriptions. There were numerous conversations among members of the research group (the authors of this chapter) during which we engaged in the iterative process of tacking among the goals we had for the design and use of EASE-C, the knowledge base regarding the challenges of comprehension instruction (RT and QtA in particular), and the video corpus. (The goals are presented in Table 18.1, and the challenges specific to RT and QtA are re-represented in Table 18.2.) These goals and challenges played an important role in both shaping the coding scheme that we eventually developed and drove the selection of clips and interviews for the video corpus.

As the general categories of codes emerged, we examined the tapes for the purpose of not only applying these general codes but also for identifying specific topics that were appropriately clustered under those codes. This process led to the set of codes and subtopics that are identified in Table 18.3. These codes in turn became the themes by which the user could search the video corpus.

With this coding scheme in place (and topics emerging), the lessons were then re-represented as clips (brief segments of 30 seconds to several minutes). Each clip was coded using one of the 10 themes and assigned topic descriptors as well. So, for example, a segment that demonstrated the teacher talking about the illustrations in a text was coded as "Using text characteristics: illustrations."

Finally, each clip was assigned a rating from 1 to 3. A 1 signified that the clip was a particularly clear instance of the theme, whereas a 3 indicated that the clip was a more subtle instance. These ratings would be useful to future research in which we planned to examine teachers' selections and representations of clips; one hypothesis being that, with more expertise, teachers would see connections among clips that were, perhaps, not initially apparent.

In the next step, we identified six "crossroads" cases for each of the two instructional procedures. A crossroads case was an instructional episode that was coded using multiple codes (as many as eight); hence, these were identified as particularly rich instructional moments that would serve as excellent springboards in using the video corpus.

TABLE 18.2
Specific Challenges Particular to Questioning the Author (QTA) and Reciprocal Teaching (RT)

Challenges particular to successful enactment of QtA	Challenges particular to successful enactment of RT:
• listening to students and weaving their comments into meaningful discussion,	• supporting students to make opportunistic and flexible use of the strategies,
• crafting questions that are both supportive and open	• learning a repertoire of ways to support (scaffold) the participation and learning of struggling readers
• identifying the targeted big ideas in text.	• understanding that the dialogues develop over time and that teachers' practices need to be coordinated with this development.

The interview clips were then excerpted, matched, and linked with the instructional events to which they referred, with every effort made to capture the teachers' reflections regarding complexity in the instruction, ambivalence (or outright dismay) about a teaching move, and surprise regarding a turn of events.

The coded video segments were entered into the EASE-C system and were then accessible for a variety of purposes. To provide an overview of RT and QtA, we prepared 20-minute sequences that teachers could view to get a sense of each instructional approach, in essence to provide a gestalt before the viewer was invited to glimpse and analyze moments of the instruction. These were called the *straight stories* for RT and QtA. In the next section, we describe how we situated the use of this environment in the professional development, using it in a manner that capitalized on its affordances.

IMPLEMENTING AND INVESTIGATING EXPERIENCE ACCELERATION SUPPORT ENVIRONMENT FOR COMPREHENSION

Little (1993) provided a helpful synthesis of the features of promising professional development with teachers. She identified three key characteristics of the most promising forms of professional development. The first is that the professional development engage teachers in the pursuit of genuine questions, problems, and curiosities, offering opportunities for meaningful intellectual, social, and emotional engagement with ideas, materials, and colleagues. The

TABLE 18.3
Searchable themes and subtopics in Experience Acceleration Support Environment: Comprehension

Building a Learning Community is about:

- Communicating to students about their roles in discussions
- Communicating expectations of students
- Authorizing students to take charge of their learning
- Encouraging students to be collaborative in their learning

Cognitive Engagement is about:

- Pacing instruction and using wait time
- Encouraging effort and persistence
- Pressing students to achieve more

Messages about Reading is about:

- Supporting students to come to a deeper understanding of the purposes and processes of reading
- Communicating to children about being flexible and strategic readers
- Monitoring for sense-making

Using Text Characteristics is about:

- Using text features (like illustrations, punctuation, text organization) to support learning from text

Modeling Expert Reading is about:

- How teachers can model identifying and connecting big ideas in the text
- How teachers can model monitoring for sense making
- Modeling use of the strategies in RT

Building Background Knowledge is about:

- Supplying children with additional information to support their understanding of the text

Navigating toward Meaning is about:

- Teacher initiated efforts to support students to comprehend text (such as scaffolding, revoicing, building upon student contributions)

Towards Being a Meaning Maker is about:

- Student initiated efforts to comprehend text (such as making connections between prior knowledge and the text, identifying points of confusion, building upon one another's ideas)

Comprehension Breakdowns is about:

- The sources of comprehension difficulty (e.g., vocabulary, decoding, referent words, incomplete or inaccurate prior knowledge) that students experience

(*Continued*)

TABLE 18.3 (*Continued*)

Repair of Breakdowns is about:

- Teacher and/or student initiated efforts to restore meaning after there has been a comprehension breakdown

Planning is about:

- Clips related to the planning of the lessons

second feature is that the professional development take explicit account of the range of experiences that teachers bring with them to the professional development context, challenging the one-size-fits-all model of professional development. Finally, she argued that teachers should be positioned to generate knowledge and assess the knowledge claims made by others. Mindful of these admonitions, we sought to design experiences for the use of EASE-C that would optimize its use, advance the learning of teachers, and inform our understanding of the strengths and limitations of this environment.

EASE-C was proposed in response to an Inter Educational Research Agency Initiative call for proposals to scale up knowledge bases regarding educational interventions. Although the focus of this chapter is not on the Inter Educational Research Agency Initiative research, it will be helpful to the reader to understand how EASE-C was situated in this professional development effort. The co-principal investigators (Annemarie Sullivan Palincsar, Rand J. Spiro, and Shirley J. Magnuson) proposed comparing the processes and outcomes of enacting a professional development condition in which teachers had access to the EASE-C system (which was Web based) with a condition in which teachers had access to much of the same video (e.g., the straight stories, as well as the crossroads cases), but without the navigability of the EASE-C system. Toward this end, the teachers in each group participated in 4 days of professional development during the summer and 2 days of professional development during the school year. The first 2 days for each group were identical and engaged the teachers in conversations about their own instructional practices as well as in the process of analyzing text for the purpose of identifying the learning goals for using that text and anticipating the challenges students might experience; in essence, we modeled a process for planning text-based discussions. On Day 3 of the summer institute, the teachers were introduced to EASE-C (the activities described below), while the teachers in the traditional video case condition viewed and discussed the video excerpts regarding RT and QtA. Data have been collected regarding teachers' learning, teachers' comprehension instruction practices, and the reading achievement of their students. Our goal is to ultimately describe the differential experiences and outcomes of these two forms of professional development, one of which (the traditional video cases) is more dependent than the other (the EASE-C system) on the presence of professional development expertise, specific to these instructional models.

Activities With Experience Acceleration Support Environment for Comprehension

In Table 18.4, we list the array of activities in which the teachers in the EASE-C condition engaged using the system. The first three activities were conducted in the context of the professional development summer institute; the last two were designed as homework to be conducted in preparation for the school-year institutes. We discuss each in turn, for the purpose of illustrating our attempts to optimize the use of EASE-C and learn about its strengths and limitations.

Assembling the Story of Scaffolding. For their initial foray into the system, we asked the teachers to view a selection (of their own choosing) of the crossroads cases and select four video clips that were useful to their thinking about scaffolding—a topic nested within Toward Meaning. We chose to delimit the task to a single topic as the participants became acquainted with the system, and we chose a topic that is in many respects at the heart of pedagogy. Finally, we selected a topic for which there were multiple associated clips across both RT and QtA. The participants were encouraged to begin with the crossroads cases, because these cases permit a significant amount of the complexity of the domain to be introduced in a cognitively manageable manner (Spiro & Jehng, 1990).

The participants were encouraged to work in groups (of two or three), to come to consensus regarding their selection, to write a rationale for their selection, and to be prepared to discuss their choices with the group. The purpose of the story was left open to the teachers but could include comparing and contrasting scaffolding across the two forms of instruction, exploring different ways of scaffolding as a function of student differences or text demands, and identifying new learning about scaffolding.

Assembling a Story Using the "Navigating Toward Meaning-Making" Theme. With the second activity, the teachers were encouraged to work at the level of theme rather than topic. This particular theme primarily features the practices of the teachers in supporting students to comprehend text; hence, this activity was designed to provide a forum for the participants to continue to develop a common language for describing practice. In addition, in a relatively short amount of time the participants would see and describe a broad array of

TABLE 18.4
Array and Sequence of Activities Using Experience Acceleration Support Environment: Comprehension (EASE-C)

- Assemble the story of Scaffolding (4 clips)
- Assemble the story with the theme [Navigating] Toward Meaning Making or cognitive engagement
- Apply special effects of EASE-C (look again, spotlighting, and color washes) to a case
- Personalized Advanced Theme Searches: Using clips to investigate a particular aspect of reading comprehension and sequencing them to “tell a story” as a visual essay (theme defined by teachers working in pairs)
- Identify a critical event in one's own teaching and plan a visual essay with EASE-C cases that would:
  Compare and contrast with critical event
  Promote a deeper understanding of the event
  Help represent the event to others
  Provide alternative response

practices in which the teachers engaged. This opportunity was consistent with one goal of the professional development, which was to communicate the need for a repertoire of practices, rather than a particular—defined—set of practices when engaged in complex comprehension instruction. Finally, teachers were encouraged to bring a critical lens to their viewing of the teachers’ practices, noting different choices the teachers might have made or questions that the teachers’ practices elicited for them, the viewers.

Applying Special Effects. Recall that one of the goals of using an EASE system is to encourage the viewer to develop the practice of looking again, with the assumption that revisiting an event or clip has the potential to lead to new observations, insights, or both. This activity was introduced by showing the teachers how special features could be added to EASE-C; for example, in one clip we added a spotlight to call attention to a particular student’s engagement; for all intents and purposes, this student appears on the video clip to be attending to everything other than the discussion. However, on looking again, when the teacher asks the group to look back at the text for evidence regarding a claim one student is making about the information, this youngster can be seen using his finger to point to that information, causing the viewer to reconsider this student’s engagement. Another example, consistent with one of our goals, was illustrating the ways in which the unfolding of a text-based discussion reflects the interactions of the text, the students, and the context in which the discussion occurs. Using color washes, we demonstrated how these three features shaped the conversation. Finally, we demonstrated how playing the clips in slow motion enabled the viewer to attend to dynamics that might well be missed in a typical viewing (e.g., the mirroring, by one girl, of the teacher's hand motions as she signaled putting together ideas in the text). After viewing these illustrations, the participants were encouraged to select clips to view repeatedly, for the purpose of determining what they noticed as they revisited events and what feature they might add to help others engage in this same noticing.

Personalized Complex Theme Searches. With this activity, which teachers were asked to complete in preparation for the first school-year institute, the teachers were encouraged to construct their own visual essay, drawing across the themes regarding any aspect of reading comprehension. For example, the teachers could focus on the causes of comprehension breakdowns, the development of comprehension over time, or the role of text structures and features as impediments and supports to comprehension. The clips teachers selected were to be entered into a personalized index, with space for the teachers to record their reasons for choosing these particular clips. Teachers then shared these essays in the institute.

Identifying a Critical Event in One's Own Teaching and Associated Clips. In the summer institute, the teachers had been introduced to the idea of critical events in teaching. For this activity, they were asked to describe a critical event in their own comprehension instruction (i.e., an event that was particularly salient because it was emblematic of their teaching; an event that was puzzling to them; an event that marked a milestone; or an event during which they achieved insight regarding a child, comprehension, or their own teaching). Using EASE-C, the teachers were asked to select video segments that were relevant to this event. They could be relevant for a broad range of reasons; for example, they were interesting contrasts, they helped the viewer to achieve an insight, or they supported hypothesizing about the event. Once again, teachers were asked to share their essays with their colleagues.

CONCLUSION

The focus of this chapter has not been on the results of using EASE-C in professional development with teachers; data analyses regarding pre- and postassessments of teacher learning, pre- and postassessments of their students' reading achievement, the conversations in which teachers engaged while using this tool, and changes in teachers' practices were all underway as we prepared this chapter. Instead, our purpose for writing this chapter has been to illustrate the scholarship that is associated with translating knowledge bases (in this case, regarding comprehension and two approaches to comprehension instruction) in the process of constructing a technological tool designed to support professional development efforts.

As the reader will note, we drew on a rich—and robust—knowledge base in the areas of text comprehension and reading comprehension instruction, in addition to a well-defined and empirically supported theory (CFT) that suggests the importance of providing learning occasions that will mitigate against such phenomena as settling for single explanations for the causes of children's challenges with text or ignoring the trade-offs that are endemic to teaching. In addition, we were influenced by the research base on teachers' learning to identify desirable features of the context in which we would implement this tool.

Thus far, our learnings from this development work suggest that the process of bringing knowledge to scale is suitably conceived of as one that depends on working in interdisciplinary ways, bringing together, for example, educators, psychologists, and technologists. The anecdotal evidence is that teachers generally enjoyed using EASE-C and reported that it was productive in supporting their learning. In addition, we have begun investigating the video data of teachers working in small groups with the EASE-C system. The teachers revealed, even in the short amount of time available to them, evidence of increased sensitivity to context and a greater appreciation for the nuanced nature of the application of practices such as scaffolding. For example, teachers were observed shifting over time from searching for a singular and prescriptive definition of scaffolding to a more situation-sensitive analysis of the forms and purposes for scaffolding.

Although these are encouraging results, the technological infrastructure requisite to using this tool in their home and school contexts was not readily available (compared with the professional development/university setting), limiting some teachers' use to the professional development context; hence, it is incumbent on technologists and school personnel to keep one another informed about the state of technology versus the state of the art in local contexts.

The time and expertise that were required to design EASE-C suggest the importance of a research agenda to establish the efficacy of these tools in terms of cost–benefit analyses. Perhaps most important, our experiences with this development work have suggested the importance of being judicious about which knowledge bases in education are ready to "bring to scale."

ACKNOWLEDGMENTS

The development of Experience Acceleration Support Environment: Comprehension, featured in this chapter, was made possible through a grant from the Inter Educational Research Agency, administered by National Science Foundation, titled "Investigating the Feasibility of Scaling Up Effective Reading Comprehension Instruction Using Innovative Video-Case-Based

Hypermedia." This grant was awarded to Annemarie Sullivan Palincsar, Rand J. Spiro, & Shirley J. Magnusson; the project officer was Anthony E. Kelly.

REFERENCES

Beck, I. L., McKeown, M. G., Hamilton, R., & Kucan, L. (1998). Getting at the meaning: How to help students unpack difficult text. *American Educator*, *22*, 66–85.

Beck, I. L., McKeown, M. G., Worty, J., Sandora, C. A., & Kucan, L. (1996). Questioning the author: A year-long classroom implementation to engage students with text. *Elementary School Journal*, *96*, 387–416.

Borko, H., Davinroy, K. H., Bliem, C. L., & Cumbo, K. B. (2000). Exploring and supporting teacher change: Two third-grade teachers' experiences in mathematics and literacy staff development project. *Elementary School Journal*, *110*, 273–306.

Bransford, J. D., & Schwartz, D. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (Vol. 24, pp. 61–100). Washington, DC: American Educational Research Association.

Bryant, D. P., Linan-Thompson, S., Ugel, N., Hamff, A., & Hougen, M. (2001). The effects of professional development for middle school general and special education teachers on implementation of reading strategies in inclusive content area classes. *Learning Disability Quarterly*, *24*, 251–264.

Buly, M., & Valencia, S. W. (2002). Below the bar: Profiles of students who fail state reading tests. *Educational Evaluation and Policy Analysis*, *24*, 219–239.

Chall, J. S. (1983). *Learning to read: The great debate*. New York: McGraw-Hill.

Chall, J. S., Jacobs, V. S., & Baldwin, L. E. (1990). *The reading crisis: Why poor children fall behind*. Cambridge, MA: Harvard University Press.

Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, *102*, 294–343.

Donahue, P. L., Voekl, K. E., Campbell, J. R., & Mazzeo, J. (1999). *NAEP 1998: Report card for the nation and the states*. Washington, DC: U.S. Department of Education.

Donovan, C. A., & Smolkin, L. B. (2001). Genre and other factors influencing teachers' book selections for science instruction. *Reading Research Quarterly*, *36*, 412–440.

Donovan, M. S., Wigdor, A. K., & Snow, C. E. (2003). *Strategic Education Research Partnership.* Washington, DC: The National Academies Press.

Duke, N. K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, *35*, 202–224.

Elmore, R. F. (1996). Getting to scale with good educational practice. *Harvard Educational Review, 66*(1), 1-26.

Fullan, M. (2000). The three stories of education reform. *Phi Delta Kappan*, *81*, 581–584.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*, 915–945.

Gersten, R., Williams, J. P., Fuchs, F., & Baker, S. (1998). Reading comprehension research for students with learning disabilities. In *Improving reading comprehension for children with disabilities: A review of research final report* (pp. 1–69), U.S. Department of Education, Office of Special Education Programs.

Goldman, S. R. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes*, *23*, 357–398.

Hewitt, J., Pedretti, E., Bencze, L., Vallaincourt, B. D., & Yoon, S. (2003). New applications for multimedia cases: Promoting reflective practice in pre-service teacher education. *Journal of Technology and Teacher Education*, *11*, 483–500.

Hiebert, E. H., & Fisher, C. W. (1990). Whole language: Three themes for the future. *Educational Leadership*, *47*, 62–64.

Hughes, J. E., Packard, B. W., & Pearson, P. D. (1997). Reading Classroom Explorer: Visiting classrooms via hypermedia. Inquiries in literacy theory and practice. In C. K. Kinzer, K. A. Hinchman, & D. J. Leu (Eds.), *The National Reading Conference yearbook* (Vol. 46, pp. 494–506). Chicago: National Reading Conference.

Lampert, M., & Ball, D. L. (1998). *Teaching, multimedia, and mathematics: Investigations of real practice*. New York: Teachers College Press.

Little, J. W. (1993). Teachers' Professional Development in a Climate of Educational Reform. *Educational Evaluation and Policy Analysis*, *15*, 129–151.

Lundeberg, M. A., Levin, B. B., & Harrington, H. L. (2000). *Who learns what from cases and how? The research base for teaching and learning with cases*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McNamara, D., Floyd, R., Best, R., & Louwerse, M. (2004). World knowledge driving young readers' comprehension difficulties. In Y. B. Yasmin, W. A. Sandoval, N. Enyedy, A. S. Nixon, & F. Herrera (Eds.), *Proceedings of the Sixth International Conference of the Learning Sciences: Embracing diversity in the learning sciences* (pp. 326–333). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Meyer, J., & Rowan, B. (1978). The structure of educational organizations. In J. Meyer and W. R. Scott (Eds.), Organizational environments: Ritual and rationality (pp. 71–96). Beverly Hills, CA: Sage.

Miller, K. F., Kelly, M. K., & Zhou, X. (2004). Learning mathematics in China and the United States: Cross-cultural insights into the nature and course of mathematical development. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 163–178). New York: Psychology Press.

National Assessment of Educational Progress. (2003). *The nation's report card: Results of the 2005 National and State Mathematics and Reading Assessments for Grades 4 and 8.*

National Reading Panel. (2000). *Teaching children to read*. Washington, DC: National Institute for Child Health and Development.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *2*, 117–175.

Palincsar, A. S., & Brown, A. L. (1989). Classroom dialogues to promote self-regulated comprehension. In J. Brophy (Ed.), *Advances in research on teaching* (Vol. 1, pp. 35–71). New York: JAI.

Palincsar, A. S., Dalton, B., Magnusson, S., Hapgood, S., Proctor. P., & Khasnabis, D. (2004, December). *Investigating verbal protocols for what they reveal about upper elementary students' text processing across narrative and informational texts*. Paper presented at the annual meeting of the National Reading Conference, San Antonio, TX.

Palincsar, A. S., Stevens, D. D., & Gavelek, J. R. (1989). Collaborating with teachers in the interest of student collaboration. *International Journal of Educational Research*, *13*, 41–54.

Pressley, M. (2000). Comprehension instruction in elementary school: A quarter-century of research progress. In B. M. Taylor, F. F. Graves, & P. van den Broek (Eds.), *Reading for meaning: Fostering comprehension in the middle grades* (pp. 32–51). Newark, DE: International Reading Association.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, *29*, 4–15.

RAND Reading Study Group. (2002). *Reading for understanding*. Santa Monica, CA: RAND.

Richardson, V., & Kyle, R. S. (1999). Learning from videocases. In M. Lundeberg, B. Levin, & H. Harrington (Eds.), *Who learns what from cases and how?* (pp. 121–136). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of research. *Review of Educational Research*, *64*, 479–530.

Sandora, C. A. (1994). *A comparison of two discussion techniques: Great Books (post-reading) and Questioning the Author (on-line) on students' comprehension and interpretation of narrative texts.* Unpublished doctoral dissertation, University of Pittsburgh, PA.

Sherin, M. G., & Han, S. (2004). Teacher learning in the context of a video club. *Teaching and Teacher Education*, *20*, 163–183.

Spiro, R. J., Collins, B. P., Thota, J. J., & Feltovich, P. (in press). Cognitive flexibility theory: Hypermedia for complex learning, adaptive knowledge application, and experience acceleration. In A. Kovalchik & K. Dawson (Eds.), *Educational technology: An encyclopedia*. Santa Barbara, CA: ABC-CLIO.

Spiro, R. J., Coulson, R. L., Feltovich, P. J., & Anderson, D. K. (1994). Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. In R. B. Ruddell & M. R. Ruddell (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 602–616). Newark, DE: International Reading Association. (Original work published 1988)

Spiro, R. J., Coulson, R. L., Feltovich, P. J., & Anderson, D. K. (1994). Cognitive flexibility theory: Complex knowledge acquisition in ill-structured domains. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 602–615). Newark, DE: International Reading Association.

Spiro, R. J., Feltovich, P. J., & Coulson, R. L. (1996). Two epistemic world-views: Prefigurative schemas and learning in complex domains. *Applied Cognitive Psychology*, *10*, S51–S61.

Spiro, R. J., Feltovich, P. J., Jacobson, M., & Coulson, R. L. (1991). Cognitive flexibility, constructivism and hypertext: Advanced knowledge acquisition in ill-structured domains. *Educational Technology*, *31*(5), 24–33.

Spiro, R. J., & Jehng, J. C. (1990). Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversal of complex subject matter. In D. Nix & R. J. Spiro (Eds.), *Cognition, education, and multimedia: Explorations in high technology* (pp. 163–205). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Spiro, R. J., & Taylor, B. M. (1987). On investigating children's transition from narrative to expository discourse: The multidimensional nature of psychological text classification. In J. Tierney, P. L. Anders, & J. N. Mitchell (Eds.), *Understanding readers' understanding* (pp. 77–93). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Snow, C. E., Burns, S., & Griffin, P. (1988). *Preventing reading difficulties in young children*. Washington, DC: The National Academies Press.

Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research*, *68*, 277–321.

Taylor, B. M., Pearson, P. D., Peterson, D., & Rodriguez, M. C. (2002). Looking inside classrooms: Reflecting on the "how" as well as the "what" in effective reading instruction. The Reading Teacher. 56 (3), 270–279.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia*. Cambridge, MA: Harvard University Press.

Vaughn, S., Moody, S. W., & Schumm, J. S. (1998). Broken promises: Reading instruction in the resource room. *Exceptional Children*, *64*, 211–225.

Wang, J., & Hartley, K. (2003). Video technology as a support for teacher education reform. *Journal of Technology and Teacher Education*, *11*, 105–138.

Wilson, V. L., & Rupley, W. H. (1997). A structural equation model for reading comprehension based on background, phonemic, and strategy knowledge. S*cientific Studies of Reading*, *1*, 45–63.

V

Conclusion

19

The 4-Pronged Comprehension Strategy Framework

Danielle S. McNamara, Yasuhiro Ozuru, Rachel Best, and Tenaha O'Reilly
University of Memphis

This chapter organizes the various strategies described in this volume within a single framework and is aligned with the COLLEGE BOARD READING STANDARDS. This chapter proposes a 4-pronged comprehension strategy framework that consists of Monitoring Comprehension and Reading Strategies in the center of the framework and four categories of strategies that comprise the prongs of the framework, including: (a) Preparing to Read; (b) Interpreting Words, Sentences and Ideas in Text; (c) Going Beyond the Text; and (d) Organizing, Restructuring, and Synthesizing Information in the Text. This chapter describes empirically and theoretically motivated strategies within each of these categories.

In this chapter, we propose a framework for classifying reading strategies and describes empirically supported strategies that fall within that framework. The chapters in this volume describe reading strategies and interventions that are theoretically motivated and have been shown to effectively improve reading comprehension. This chapter organizes the various strategies described in this volume within a single framework. This framework also emerged from a collaboration with the College Board to revise their English, Language, and Arts Standards for middle school through college-ready high school students. The College Board Reading Comprehension Standards describe the process by which successful readers construct the meaning of texts. The first three standards are (a) Comprehension of Words, Sentences, and Components of Texts; (b) Using Prior Knowledge, Context, and Understanding of Language to Comprehend and Elaborate the Meaning of Texts; and (c) Author's Purpose, Audience, and Craft. Most important, the fourth reading standard included within the College Board standards document is Using Strategies to Comprehend Texts. As far as we know, the College Board's is the first standards document to include reading strategies as a separate standard. This was done in recognition of the growing empirical evidence for the importance of reading strategies for successful reading. Much of that evidence is summarized in the chapters of this volume.

The reading comprehension strategies framework that emerged during the College Board project is supported empirically and also aligned with theories of the reading comprehension process (e.g., chap. 1, this volume). The first underlying notion of this framework is that the reader's mental representation of a text consists of multiple levels of comprehension, and the principal levels are the textbase and the situation model (e.g., Kintsch, 1998). The *textbase* is the reader's understanding of the words and sentences that are explicitly presented in the text. The *situation model* is the reader's understanding that goes beyond the text, that is, an integration of the text with what the reader already knows about the world and the topic of the text.

The second theoretical notion that drives this framework is that successful readers construct coherent mental representations (i.e., situation models) from text by actively processing and integrating concepts from the text and related concepts (e.g., Graesser, Singer, & Trabasso, 1994). Critical to comprehension is the activation of domain and general world knowledge that helps the reader to develop a more coherent mental representation of the content of the text. Also critical is making links among ideas, synthesizing ideas, and organizing concepts into a global understanding. These processes require that the reader be metacognitively aware while reading, monitoring the success of the reading process and the success of the reading strategies used.

These notions of text comprehension inspired the Center and the 4-prongs of reading strategies comprising the four-pronged framework (see Figure 19.1). The center of the 4-pronged framework is Monitoring Comprehension and Reading Strategies. Comprehension monitoring is assumed to be intrinsically linked to the reader's use of the other four categories of strategies. The first category of reading strategies is Preparing to Read. This category includes setting or recognizing the goals of reading and using prereading strategies to guide the reading process. The second category includes strategies to Interpret Words, Sentences, and Ideas in Text. These are strategies that help the reader focus on and understand the words in the text and, in essence, to develop a coherent textbase level of understanding. The third category includes strategies to help the

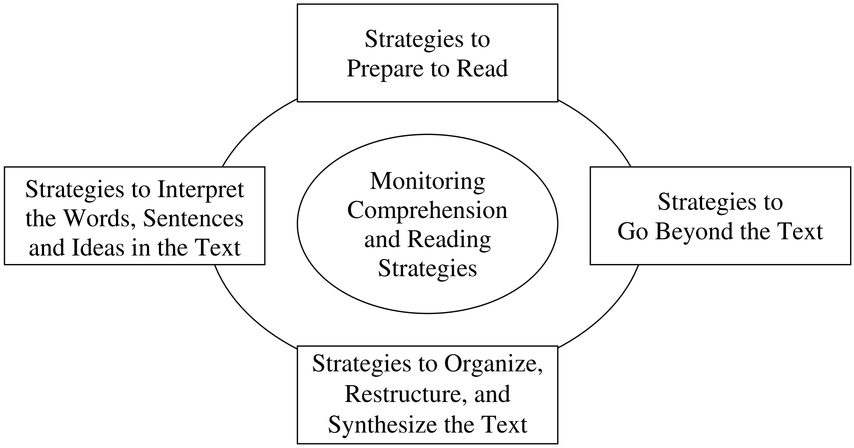


Figure 19.1. The 4-pronged framework for reading comprehension strategies.

reader Go Beyond the Text by connecting the text with their prior knowledge. The fourth category includes strategies that help the reader Organize, Restructure, and Synthesize the information in the text. Hence, this framework is based on the general premise that reading strategy use is intrinsically metacognitive, with the monitoring of the comprehension at its core. Metacognitive reading strategies induce and support the use of monitoring, which in turn facilitates the use of various reading strategies.

Unlike others, this reading strategy framework does not explicitly follow from the notion that there are strategies that are used *before, during, and after* the reading process (e.g., Saricoban, 2002; Schmitt, 1990). Although each of the categories could be placed more or less before, during, or after reading, this notion was abandoned relatively early in the development of the framework because there are so many strategies that take place at multiple stages of the reading process. Although it is clear that preparing to read takes place before reading, it could be argued that organizing, restructuring, and synthesizing could take place while reading as well as after reading. Moreover, the process of monitoring comprehension clearly takes place at all three stages in the reading process.

It is further notable that this reading strategy framework addresses strategies for comprehending text, not decoding text. Just as this volume is concerned with the comprehension processes and not decoding processes, so too is this framework. Exclusion of the decoding process from the framework is based on the assumption that understanding text consists of decoding and comprehension processes and that decoding and comprehension are relatively separable processes (e.g., Cain, Oakhill, Barnes, & Bryant, 2001). Our focus here is on strategies for comprehension; thus, what the reader will not find in this chapter are strategies to decode text.

Also not included in this chapter are instructional techniques. This chapter focuses solely on strategies that students can be taught, with the goal of eventually using the strategies on their own, and does not focus on which kinds of instructional techniques can best be used to teach the strategies. Thus, unlike the other chapters in this volume, this chapter does not discuss techniques to teach students to use these strategies, such as peer tutoring (chap. 7, this volume), Reciprocal Teaching (chap. 18, this volume), iSTART (chap. 16, this volume), or Concept-Oriented Reading Instruction (chap. 10, this volume). The components or individual strategies that comprise these types of programs, at least the empirically supported ones, are described in this chapter, but the programs themselves are not, because these programs generally include multiple reading strategies. However, we would like to point out that the instructional techniques used to teach reading strategies play a critical role in helping students use reading strategies effectively and decide which strategies to use and when. For example, in chapter 7, regarding peer-assisted learning, Fuchs and Fuchs discuss how peer tutoring facilitates the use of various reading strategies (e.g., question generation and summarization) among children at the elementary level. Furthermore, McNamara et al.'s chapter on iSTART illustrates when self-explanation training is more likely to support the use of text-based strategies (e.g., paraphrasing) versus strategies for going beyond the text (e.g., elaboration) among high school students.

THE COMPREHENSION CHALLENGE

Comprehending text is clearly challenging. First, it involves a series of subprocesses, such as word decoding, lexical access, syntactic processing, and higher level inferencing based on the discourse context (Vellutino, 2003). Second, readers must successfully execute these processes in a coordinated fashion in order to successfully comprehend a text. Many readers struggle at one stage or another in the cognitive processes involved in reading comprehension. The stage(s) at which readers struggle may differ depending on the reader and text characteristics. For example, experienced readers may encounter problems when reading unfamiliar text if they lack the specific domain knowledge required to make certain inferences (e.g., O'Reilly & McNamara, in press; Shapiro, 2004). On the other hand, inexperienced readers may encounter more problems with vocabulary, syntactic processing (e.g., a long sentence with multiple relative clauses), or with integrating information (e.g., understanding the motivation for a murder interspersed across several chapters).

Finally, to make the comprehension process more complicated and demanding, many circumstances require readers not only to understand the text as they read but also to remember the critical content of the text sometime later in various situations (e.g., talking about the content, answering questions about the content, etc.). In a sense, the purpose of reading is often to learn or remember something about the text's contents at a later time. For this reason, the reader must monitor whether text contents have been comprehended reasonably accurately or as intended by the author and whether the information learned from the texts can be used to accomplish foreseeable goals (e.g., essay writing, answering questions, participating in discussion, etc.).

The process of being aware of understanding is often referred to as *metacognition* or *metacomprehension*. Although there is some disagreement on the exact definition of metacognition (Kurtz, 1991), researchers generally agree on two critical features: (a) self-regulation and (b) the use of strategies (Schraw & Dennison, 1994). Certain strategies are more or less effective depending on the task, the content, and the context (Alexander & Judy, 1988; O'Reilly, Symons, & MacLatchy-Gaudet, 1998). Therefore, successful comprehension and learning hinges on the ability to identify which strategies are effective for specific contexts and the ability to select the optimal strategy to improve comprehension (e.g., Everson & Tobias, 1998; Maqsud, 1997; van Kraayenoord & Schneider, 1999). So, although having a repository of strategies is advantageous, the reader must also self-regulate (or monitor) the success of those strategies and the reading process. The construction of an effective model of a text requires that the reader constantly monitor the coherence of the mental representation of the text and update it when new information is provided or when new information contradicts the reader's existing mental model.

Monitoring understanding and learning is, however, a challenging and sometimes impossible task for many readers. The literature suggests that people's ability to monitor comprehension is relatively poor: Readers are often poor judges of their own levels of comprehension and learning (e.g., Baker, 1985; Cavanaugh & Perlmutter, 1982; Markman, 1977; Pressley & Ghatala, 1990). There are many reasons why this might be the case. One reason people fail to monitor their performance while reading is because some readers appear to have a tendency to judge comprehension success at relatively low or local levels of processing (e.g., McKoon & Ratcliff, 1992). This low level of processing generally corresponds to focusing on the processes of word decoding and forming the meaning of the words into sentence-level comprehension. As a consequence, readers often pay little attention to whether the text makes sense at the global level (e.g., coherence relations among multiple sentences) or whether the overall text content is consistent with their pre-existing beliefs or understanding of the subject matter. Thus, when the text makes sense at the local level, readers often develop an illusion of understanding (Glenberg, Wilkinson, & Epstein, 1982), one that does not accurately reflect the potentially fragmented nature of the reader's mental representation.

Another potential explanation for the difficulty of comprehension monitoring is that people monitor their comprehension in an inappropriate way. In particular, there are reasons to believe that circumstances of the reading situation lead people to overestimate their level of comprehension and learning from the text. That is, when people read and then subsequently test their understanding of the contents, the critical information being assessed is often available to the reader, if only in working memory. As such, the reader is under the illusion that he or she can reconstruct the information at a later point in time (Thiede, Anderson, & Therriault, 2003). Overall, monitoring of comprehension is often inadequate because readers fail to take into consideration the delay between reading and assessment and the availability of the cues in the text (i.e., specific parts of the text content). Consequently, readers tend to perform poorly on subsequent comprehension tests.

THE 4-PRONGED FRAMEWORK: FACILITATING COMPREHENSION USING READING STRATEGIES

Thus far, our introduction of processes involved in reading comprehension indicates that there are a number of potential obstacles a reader may encounter when attempting to comprehend text. Fortunately, studies of reading comprehension have identified techniques that help poor readers overcome various difficulties associated with comprehension. We refer to these techniques as *strategies*. Strategies, unlike skills, are conscious and generally effortful. Strategies are also purposeful. Some have quite specific purposes or goals. For example, readers may use strategies to become efficient at monitoring their level of comprehension, or to process the meaning of a particular sentence. This can occur by consciously engaging in particular types of activities such as analogically connecting the sentence content to general knowledge or personal experience and/or generating specific questions about the sentence meaning and checking whether one can provide a sufficient answer to that question. Other strategies are more general in the sense that they help the reader become more active and engaged in the reading process (e.g., participating in question asking), which in turn makes the reading activity more meaningful.

One factor that distinguishes successful from less successful readers is the use of reading strategies, particularly when comprehension problems are encountered (Brown, 1982; Long, Oppy, & Seely, 1994; Oakhill, 1984; Oakhill &

Yuill, 1996). Also, as the chapters in this volume have shown, providing instruction to use reading strategies that help readers process text more actively and deeply in turn helps less skilled readers to more successfully understand text. In this concluding chapter, we describe these key strategies in terms of their contribution to reading comprehension processes.

In the following sections, we present reading strategies that help the reader overcome various levels of comprehension difficulties under the rubric of the four-pronged framework. We start by discussing the notion of comprehension monitoring, because it is a critical process that affects comprehension at all levels. The subsequent sections focus on reading strategies that relate to different aspects of comprehension processes, including preparing to read; interpreting the words, sentences, and ideas; going beyond the text; and organizing, restructuring, and synthesizing. Although each strategy is generally listed under one category, the distinction is primarily for organizational purposes. It is important to note that some strategies are more general in the sense that they are likely to facilitate multiple cognitive processes underlying reading comprehension and thus are more likely to cross boundaries within our rubric.

Monitoring Comprehension and Reading Strategies

- The reader monitors comprehension by engaging in strategies to determine the level of understanding. Strategies include generating questions while reading to assess understanding and noting or marking where and when comprehension fails.
- The reader uses strategies to assess postreading comprehension, memory, and learning. Strategies include generating key words, answering questions, taking a test, or attempting to recall the text after a delay following reading.
- The reader adjusts reading strategies to improve comprehension.

Comprehension monitoring plays an important role in comprehension at all levels and stages of reading. Before reading, learners can think about what they know about the topic, the text structure, and genre and make predictions about what they think will be covered. During reading, learners can check to see if predictions and their prior knowledge are consistent with the text. The reader must constantly monitor whether he or she understands the text at various levels. Word meanings must be accessed; if they are unfamiliar, action must be taken to understand word meanings. Meaning must be derived from sentences and paragraphs to create a coherent picture of what is discussed in the text. At any one of these levels, comprehension can fail. However, the critical point is that when comprehension fails, the reader must recognize the failure and take action toward rectifying the coherence break.

Thus, monitoring both comprehension and, by consequence, the success of reading strategies, is necessary for skilled, successful reading. At one level, comprehension monitoring falls out of using reading strategies; that is, the use of a reading strategy requires that the reader be cognizant, at some level, of the comprehension process. Thus, whenever he or she is using reading strategies, the reader must to some extent be monitoring comprehension. It is important, nonetheless, that the reader understand that the center of comprehension is the recognition of the importance of comprehending what is read—that the goal of reading is to create coherence from the text. This is the essence of comprehension monitoring.

The ultimate aim is to foster comprehension monitoring by overtly engaging in activities that support cognitive control and self-regulation. Many of the strategies discussed in the following sections serve the role of facilitating a specific aspect of comprehension; however, many of these strategies also serve a secondary role by facilitating the reader's monitoring of understanding. For instance, defining a goal before reading can serve as a benchmark for judging whether the reader's standards for comprehension are satisfied. Having a specific purpose can guide understanding by providing a means to signal when comprehension and reader goals are misaligned. Similarly, previewing sections of the text before reading can function as setting a template for conceptualizing what is already known and, consequently, what requires further clarification when the text is read in detail.

There are also specific strategies that readers can use to facilitate and induce comprehension monitoring. A common way to induce monitoring is to mark the text or take notes while reading. This strategy is further elaborated in this framework under the rubric of strategies to interpret meanings of words, sentences, and ideas; however, annotating text is also directly relevant to comprehension monitoring. Marking where comprehension breaks down helps the reader to remember where to return to reread and further process the text. Taking notes helps to maintain attention, externally records the reader's understanding, and potentially provides a summary of the main points from the text.

One of the most effective ways to directly foster comprehension monitoring is for readers to test their understanding of the text. According to Pressley and Ghatala (1990), readers are better able to judge their understanding when they are tested on whether they know the material. Consciously monitoring comprehension through self-testing helps to offset the potential negative effects of the automatic nature of reading, or minimalist processing (McKoon & Ratcliff, 1992). It is important to re-emphasize, however, that the effects of testing comprehension on monitoring are influenced by the amount of delay between reading and testing (Thiede et al., 2003). If the text content is fresh in the mind of the reader, then the reader is less likely to recognize comprehension failures.

In contrast, after a delay the reader can more readily judge whether the material was understood and retained.

One method of improving monitoring that has some empirical evidence for its effectiveness is for the reader to generate key words from the text after a delay (Thiede et al., 2003). Another technique that allows the reader to test comprehension is for the reader to generate questions about the author's intentions and the text content (King, 1994; see Rosenshine, Meister, & Chapman, 1996, for a review). These questions can be generated at various points: while reading, after reading, and after a delay. Questions generated while reading can be used to assess comprehension of the content at various points in the text. Questions generated after reading may serve as a means to review, or ensure that the reader has both understood and remembered the content. Attempting to recall or summarize the text can also serve as a means of self-testing comprehension. If the reader is not able to report the gist of the material, then it is likely that the reader has only a fragile understanding of the text, and more processing is needed. Like generating questions, recall and summarization are strategies that can be used at various points in the reading process: while reading, after reading, and after a delay.

In summary, all of these strategies serve the purpose of affording readers the opportunity to explicitly check their understanding while reading. This can be achieved by making notes while reading or by directly testing comprehension. Techniques for testing comprehension include generating key words from the text, generating questions about the text, answering questions about the text, recalling the text, and summarizing the text. There is some evidence that testing comprehension after a delay is most effective, because the information is not fresh in the reader's mind. The reader can better judge then whether the information will be retrievable. Testing after a delay also allows the reader to practice the ultimate goal: remembering the information from the text at a later time.

Preparing to Read

- The reader identifies the purposes and goals of reading.
- The reader uses prereading strategies, such as previewing sections of the text, creating concept lists and maps, and generating prereading questions before reading.
- The reader uses the information from these prereading strategies to guide the reading process and to select appropriate reading strategies.

A critical first step to successful deep-level comprehension is reading preparation. Most of the time, understanding text requires more than visually tracing words or sentences in texts. Readers need to actively think about the words, sentences, and paragraphs according to some type of personal and situation-specific criteria (van den Broek, Lorch, Linderholm, & Gustafson, 2001). In this section, we discuss two important components of preparing for active reading. The first regards the importance of having goals for reading, whereas the second concerns the use of prereading strategies.

Setting a Reading Goal. A logical starting point for successful comprehension is for the reader to have goals and purposes for reading (e.g., for entertainment, to answer general and specific questions, to learn new information, to solve a problem). Setting a goal or purpose gives the reader an idea about how to be selective in the reading of material and to focus on the critical content (Pressley, 2000). Having explicit overarching goals is important for the reading process, because the reading goal serves as the foundation for comprehension and learning criteria against which readers monitor the progress of their reading activity. For example, readers' comprehension criteria for a paragraph describing a key cause of the American Revolution may differ when a passage is being read for general entertainment compared with when it is read to write an essay on the topic. A number of empirical studies indicate that different reading goals are associated with different levels of comprehension (e.g., van den Broek et al., 2001). In a sense, having no explicit goal for reading would reduce comprehension criteria to a very low level (e.g. minimalist processing; McKoon & Ratcliff, 1992). When approaching the reading task in this manner, readers often do not maintain coherence much beyond adjacent sentences, reducing the likelihood that the reader attains a global understanding of the text.

Reading goals vary from fairly general (e.g., entertainment vs. learning) to very specific (e.g., preparing to answer questions about a given topic, or understanding specific information about the nature of compound molecules). The nature of the goal and the specificity of the goal will depend greatly on the nature of text being read but will also be affected by the level of the reader's pre-existing knowledge relating to the text contents as well as the reader's comprehension abilities.

Whereas good comprehenders may be able to set very specific goals and flexibly shift goals during the reading process, poor comprehenders are unlikely to make such finely grained adjustments (Pressley, 2002). Similarly, whereas readers with a lot of topic knowledge can set very specific goals based on their expectations of text contents, less knowledgeable readers may not be able to set specific goals. Thus, it is impossible to prescribe what an optimal type of goal should be for all readers; each reader needs to learn how to adjust reading goals through experience. Nonetheless, as a general guideline a reader should move from having general goals to specific goals as he or she becomes a more proficient reader and as he or she moves through a text. Thus, reading training should begin by encouraging readers to have some goals, even for lower level students. Training should then focus on the nature of specific goals; students eventually should learn to have goals appropriate for varying text genres (e.g., narrative vs. expository) and reading situations (e.g., information search, answering questions, writing critical essays). As readers become more proficient in reading, and in using reading strategies, they will also become more proficient at adjusting their reading processes and strategies to match their goals.

Using Prereading Strategies to Guide the Reading Process. Prereading strategies serve a number of purposes, including helping readers set goals to guide the reading process and activating prior knowledge. Prereading strategies include techniques such as previewing sections of the text (scanning or skimming the text), creating concept lists and maps before reading, and generating prereading questions.

Previewing the text is possibility the most well-known prereading strategy. Previewing strategies involve surveying the text before reading. This generally involves reading over key parts of a text, such as the title, subheadings, bold or italicized words, figures and tables, the introduction and conclusion, and key sentences (e.g., the first sentence of each paragraph). This is sometimes referred to as *scanning the text*. *Skimming* the text is similar but involves reading more of the text, quickly.

There are several ways previewing helps readers set up goals that guide comprehension processes. First, previewing is helpful for allowing readers to become familiar with text contents and activate prior knowledge. Second, previewing helps readers identify and take advantage of the text structure (see chaps. 8 and 14, this volume). Reading over key parts of a text, such as the title, subheadings, and bold or italicized words, helps readers to better understand some of the main themes or questions associated with the text. As such, previewing is particularly beneficial when reading expository materials, because these texts often have explicitly marked structures (e.g., introduction, supporting details, conclusions, etc.) that help readers identify the goals of the reading. Research has indicated that previewing leads to comprehension gains for students from the fourth grade through middle and high school levels (Richardson & Morgan, 2000; Spires, Gallini, & Riggsbee, 1992).

The reader can perhaps increase the effectiveness of previewing or prereading by generating a concept list or map based on the title, subtitle, and some important concepts contained in the text (e.g., highlighted words contained in informational materials). *Concept maps*, which can be produced before, during, or after reading, require the reader to transform texts into a graphic representation to indicate the links between key ideas. Concept maps oftentimes consist of concepts placed in graphical-geometric shapes that are arranged in hierarchical order (e.g., content concepts represented in ellipses and the rhetoric structure concepts represented in rectangles). When used before reading, concept maps are conceptualized as a special advanced organizer that the reader may use to explore connections between information cited in major headings and subtitles. The process of concept mapping has been regarded as a mind tool (Jonnasen, 2000). First, the reader can offload information (i.e., reduce working memory load) to invest more cognitive resources in the comprehension processes, thus leading to more meaningful learning (Novak, 2004). Second, the process of creating a concept map allows the reader to actively engage with a text and construct a mental model of the text contents (Kozminsky, 1992). As we discuss later, in the section on organizing, synthesizing, and restructuring text content, creating concept maps help readers become aware of the complex relations between key concepts that are not always readily discernible from the linear structure of the text.

A recurrent strategy across the categories of this framework is generating questions. Likewise, generating prereading questions helps readers to better formulate their reading goals and to activate relevant prior knowledge about the text. Question asking often builds on previewing techniques. As we discussed with reference to importance of comprehension monitoring, having questions and seeking answers to the questions while reading helps guide the reading comprehension process. Guiding questions can be generated as a preparation process before reading. For example, readers may turn subheadings into questions, which direct their focus when reading. Readers may generate a priori questions about theories identified while previewing a science chapter. These questions allow the reader to understand why related ideas in the text make sense. Empirical studies indicate that questioning techniques play an important role in supporting reading comprehension from middle school through college levels (King, 1989; chaps. 11 and 12, this volume; Rosenshine & Meister, 1994). Indeed, a meta-analysis of reading comprehension strategies indicated that self-questioning is one of the most effective methods for improving comprehension (Haller, Child, & Walberg, 1988).

In summary, strategies to prepare a learner to read help him or her formulate the reading goal and determine what he or she already knows about the topic and, consequently, determine the way he or she needs to process information discussed in the text. Ultimately, readers should know when and why to use prereading strategies. Furthermore, readers should use information gathered from prereading practices to guide the reading process and select other strategies to use during and after reading.

Strategies to Interpret Meanings of Words, Sentences, and Ideas in Text

- The reader uses text-focused strategies such as rereading, paraphrasing, and chunking.
- The reader marks and annotates the text and takes notes.
- The reader makes intentional bridging inferences that connect back to previous sentences and ideas.
- The reader uses close reading.
- The reader uses knowledge of text structure.

Successful reading comprehension depends to a large part on whether the reader is able to construct a coherent textbase level of understanding (Kintsch, 1988). The textbase understanding is constructed from the information explicitly stated in the text. Forming a textbase representation requires efficient decoding and rapid access to word meanings and the efficient construction of sentence meaning based on the syntactic information contained in the sentence. However, many readers experience difficulty in constructing a textbase understanding, particularly when they encounter unfamiliar words and/or complex syntactic structures (Rowe, Ozuru, & McNamara, 2006). In this section, we describe five distinct strategies that help readers form a more coherent textbase understanding.

Using Text-Focused Strategies. One important strategy for comprehension concerns using text-focused strategies to make sense of critical words and sentences encountered while reading. By the end of elementary school or the start of middle school, most children will be able to use a range of strategies for processing information in the text. The simplest strategy to use when comprehension difficulties occur is to backtrack and reread the confusing segment or previously comprehended parts (Bereiter & Bird, 1985). Although rereading may not solve the comprehension problems or difficulties by itself, the processes to repair the problems cannot occur without reprocessing the part of the text that caused the comprehension problem. In this sense, rereading is an important first step to the use of more active strategies that facilitate reading comprehension.

A next step would be to paraphrase the text. *Paraphrasing* is the rewording of some portion of a text, usually a sentence or two, by using different words that are more familiar to the reader and phrased in a fashion more habitual to the reader. Paraphrasing requires transforming the surface characteristics of the sentence by replacing the content words or syntactic structure of the sentence. This induces the reader to process the information somewhat actively by accessing related but different lexical items. This process helps the reader to better understand the text (Rosenshine & Meister, 1994), albeit at a textbase level. Paraphrasing is also effective because it externalizes the reader's understanding of the information in the text, which in turn helps the reader monitor comprehension more closely (see chap. 16, this volume). Difficulties readers experience when paraphrasing a sentence often indicate the presence of a comprehension problem. Indeed, McNamara (2004) found a positive correlation between inaccurate paraphrases and poor comprehension of a text. However, if readers recognize problems while paraphrasing, they may then use additional strategies, such as generating inferences (Todaro, Magliano, Millis, & Kurby, & McNamara, 2006) or finding the meaning of words from external sources.

Another text-focused strategy that helps the reader make sense of sentences is to *chunk* words in a sentence into short, meaningful phrases (usually three to five words). The idea is that chunking helps the reader break up a long and often complex sentence into manageable sections. Empirical work indicates that chunked material separated into meaningful related groups of words improves comprehension for some readers, in particular, those classified as low-ability readers (Casteel, 1989). However, we must caution that other research indicates that excessive chunking may hinder comprehension (Keenan, 1984).

Marking and Annotating. As the reader becomes more experienced and develops competency in verbal or literacy skills, he or she may chose to make notes about specific concepts and ideas in a text while reading or after reading. As discussed earlier, annotating the text and taking notes can be key to monitoring comprehension. Note-taking is also important for increasing comprehension of and memory for information cited in the text (Wade, Trathen, & Schraw, 1990). There are different kinds of note-taking, such as citation of facts and summaries of important information and quotations. Notes may also include nonlinguistic forms, such as sketches. For note-taking to be most helpful for comprehension, the notes should be produced in a way so that they are meaningful to the readers and can be reviewed (T. H. Anderson & Armbruster, 1986). Indeed, for deeper level comprehension the reader should make notes to synthesize critical ideas rather than merely copy key concepts (V. Anderson & Hidi, 1988–1989). However, synthesizing main ideas is a more complex process that involves a more global understanding of a text (see the section on strategies to organize, restructure, and synthesize the text content).

Bridging Inferences. One of the challenges to comprehending texts is to make connections among key concepts, arguments, and theories instead of processing concepts in isolation. The ability to link concepts and ideas is especially important when one considers that many texts do not explicitly link related information (e.g., Beck, McKeown, Sinatra, & Loxterman, 1991). The process of generating *bridging inferences* plays an integral role in helping the reader build a global representation of the text (Magliano & Millis, 2003; chap. 16, this volume). Thus, it is important that the reader identifies ways of linking information from different parts of the text by making bridging inferences.

Although bridging inferences often require information from outside of the text, and thus could be categorized under the rubric of going beyond the text, they mostly involve the understanding of information that is explicit in the text and relationships between ideas that are presented in the text. Bridging inferences take different forms that incorporate varying levels of complexity, depending on the reader's ability. For example, young, low-ability, or low-knowledge readers may bridge ideas only between adjacent sentences (e.g., McNamara, 2004). More skilled or high-knowledge readers, however, will be more likely to make connections between ideas in multiple sentences or links between implied organizational patterns (Magliano & Millis, 2003; chap. 5, this volume). Empirical studies indicate that bridging inferences or text-based inferences play a critical role in the comprehension of and learning from texts (R. C. Anderson & Pearson, 1984; Cain et al., 2001). Foremost is that successful comprehenders are more likely to make inferences (e.g., Long et al., 1994; Magliano & Millis, in press). There is the question of whether the generation of connections between parts of the text is a process (skill) or a strategy. By and large, it is not truly a strategy but rather a process that is necessary for successful comprehension. Nonetheless, because it is so critical to successful comprehension, we include it here as a strategy. One important question is: How can readers make bridging inferences? One method is simply for the reader to have the intention to make bridging inferences, to know that they are necessary, and to consciously seek out the relationships between ideas in the text. Indeed, simple instruction to generate inferences can help less skilled readers better understand text (Hannon & Daneman, 1988; Yuill & Joscelyne, 1988). Other means of making inferences include generating and answering questions, thinking aloud, and explaining the text. Also, the use of any strategies that emphasize the relation or organization of the ideas in the texts, such as creating concept maps or summarizing, should facilitate the process of generating bridging inferences. We describe these strategies in more detail in subsequent sections.

Close Reading. Guided *close reading* is a common strategy in classrooms. In essence, it refers to the use of guides to closely examine the meaning of words, sentences, and paragraphs and to examine the semantic, syntactic, and stylistic nuances of language and authors' craft. This strategy, like that of generating bridging inferences, straddles the categories of text-focused strategies and strategies to go beyond the text. Because close reading is generally focused on understanding the textbase, we place it here with text-focused strategies.

Close reading is a method that involves paying especially close attention to what is printed on the page by rereading and analyzing particular parts of a text. It is a much more subtle and complex process than the term might suggest. Close reading means not only reading and understanding the meanings of the individual printed words but also making the reader sensitive to all the nuances and connotations of language as it is used by writers. Close reading can focus on a wide range of issues, including discerning a word's particular meaning or the syntactic construction of a sentence, to thematic progression, author's craft, or a view of the world that a text might offer. It involves almost everything, from the smallest linguistic items to the largest issues of literary understanding and judgment.

There are various forms of close reading (or guides), and they vary in complexity. During the elementary school years, many young readers are able to use the close reading method to pay attention to the surface linguistic elements of the text, including aspects of vocabulary, grammar, and syntax. For example, children may search for inconsistencies of syntax. Sentence transformation activities help the reader understand how complex sentences are constructed, and sentence recombination helps the reader to learn how ideas are linked together in a text (Herrell & Jordan, 2002). Close reading is also important for semantic processing of text contents, in particular exploring the meanings of words and relationships between word meanings. An in-depth understanding of the text requires the reader to attend to the information words *yield*, *connote*, and *denote*. Readers may explore word meanings by extrapolating meaning from the texts or by outside text sources (e.g., looking up meanings in dictionaries). In terms of exploring word meanings in context, readers may draw on a range of text-based information (Penno, Wilkinson, & Moore, 2002; Sternberg & Powell, 1983), including morpheme-based cues (e.g., prefixes and suffixes) and meanings of surrounding words. Readers may also infer word meanings by interpreting cues contained in surrounding sentences and paragraphs (e.g., information about temporality or causality).

As students become more competent readers, they may use close reading to interpret meanings implied by the text or to analyze the author's craft. For example, the reader may use close reading to search for internal consistencies and inconsistencies among concepts, ideas, and theories or to explore the logical development of ideas in relation to the author's perspective (Baker, 1985). Alternatively, the reader may use close reading to note the relationship of elements of the text to things outside it, such as pieces of writing by the same author or other writings of the same type by different authors.

Close reading is as much an instructional technique as it is a strategy. First, learning how to engage in close reading emerges through instruction and the use of guides. Also, close reading generally encompasses or requires various strategies (e.g., bridging, questioning the author, elaborating, rereading, referring to external sources). Thus, close reading is a specific term referring to a mixture and synthesis of the other more specific strategies discussed elsewhere in this chapter. Nonetheless, because it is a common strategy, and because it is generally effective, we have included it here as a strategy that helps the reader better understand the text and to understand the implied meanings of the text.

Using Text Structure. The fifth textbase strategy is to use general knowledge about the structure of texts to make inferences about the text. The reader can make connections among ideas by attending to the manner in which a text is structured and organized. As more explicitly outlined in chapters 8 and 14 of this volume, it is helpful for the reader to understand the role of text structure, because the structure helps the reader organize the content and because it contributes to the building of a mental representation of text content.

The way texts are structured varies as a function of text genre. Broadly speaking, texts can be classified as *narrative* or *expository*. During the early school years, children encounter narrative texts that follow a story grammar (e.g., descriptions of setting, characters, and other attributes of the story, alignment of causal relation and temporal relations). Also, there is an argument that the narrative format of a story is a primary way of organizing everyday events for most people (e.g., Bruner, 1986). As a result, many children understand narrative text structure (e.g., Williams, Hall, & Lauer, 2004), or it can be easily taught by asking students to identify the main story elements (Williams, Brown, Silverstein, & deCani, 1994; Williams et al., 2002). Knowing that narratives follow a sequential line of events, the reader is in a good position to process contents related to the story or predict how the story may unfold without previewing the text content beforehand.

Around the third or fourth grade, children are exposed to an increasing number of expository materials, such as science and history books. Expository text differs from narrative text in many ways. Perhaps the most salient way is that the vocabulary tends to be less familiar and the concepts more challenging. Expository structure also differs from that of narrative structure in that expository texts typically consist of a variety of abstract and logical relations (e.g., division of information into main headings and subsections) organized around a variety of discourse structures (compare and contrast, deduction or induction, argument and demonstration; Coté, Goldman, & Saul, 1998). In addition, many key concepts in informational textbooks are highlighted in boldface or italic text, which means they are important to understanding a particular topic area. Understanding the organizational structure and features of expository texts is critical for processing contents. For example, attending to subsection titles helps the reader build a more global understanding of themes and information related to the topic of interest. Furthermore, boldface or italic words draw attention to critical concepts and definitional information contained in sentences surrounding the key concepts.

Students, especially those at the elementary level, struggle to understand expository structure (Kamberelis & Bovino, 1999; Langer, 1986). It is therefore advisable that early interventions focus on teaching children about the structure of expository texts, not least because the focus of reading shifts from reading narrative texts for the purpose learning to read to expository texts during the third and fourth grades.

Awareness of text structure continues to be important throughout formal schooling. This is because students encounter new styles of text (e.g., poetry, critical reviews, and biographies) as their reading competency develops. Each style of text differs in structure, and thus an understanding of new structures is important for determining which strategies to use (e.g., elaboration vs. prediction) to aid comprehension.

Strategies That Go Beyond the Text

- The reader generates questions.
- The reader uses think-aloud and self-explanation.
- The reader uses visualization or imagery.
- The reader takes advantage of external sources of information.

There is ample evidence for the need to make use of information that goes beyond the text, including prior knowledge and experience. There are at least two reasons why the use of prior knowledge is particularly important for comprehension and learning from texts. First, virtually no text describes the intended situation, event, or information in a complete fashion; authors assume a certain amount of reader knowledge (Britton, Gulgoz, & Glynn, 1993). Thus, texts have *conceptual gaps*, or *cohesion gaps*, and the reader must often rely on prior knowledge to fill those gaps (e.g., McNamara & Kintsch, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996). Second, comprehension and learning almost by definition involve relating the text content with what one already knows relatively well. Learning is enhanced by integrating new information with prior knowledge (Kintsch, 1988, 1998), and learning suffers when newly encountered information is not integrated with pre-existing knowledge. So, strategies that are designed to capitalize on knowledge and experience improve comprehension by helping the reader form a more coherent situational model. In the following sections, we briefly describe five strategies that focus on improving comprehension through the use of knowledge and experience.

Question Generation. We have chosen to include question generation under several of the prongs in the framework, because generating questions can take on multiple forms and serve multiple purposes. Also, generating questions has been shown to be a powerful influence on improving students' comprehension (Beck & McKeown, 2001; Bradtmueller, 1983; Clark, Deshler, Schumaker, Alley, & Warmer, 1984; Inagaki & Hatano, 1974; King, 1994; King & Rosenshine, 1993; Rosenshine et al., 1996). As discussed in previous sections, readers can generate questions about the text before, during, and after they read. Questions generated before reading help activate prior knowledge and serve as guides for checking for information that the reader does not understand. Questions generated during and after reading can serve as a form of self-testing to assess online and offline comprehension. King (1989) provides an exemplary list of question stems that can be used for a wide variety of subject domains. For example, Table 1 in King (1989, p. 372; see also chap. 11, this volume) lists some possible generic question stems that readers can ask themselves while they read: "Explain why … "; How is … related to … ?"; "What do you think would happen if … ?" Question stems like these and other types of questions are beneficial because they activate relevant knowledge, support active processing, and foster comprehension monitoring. Question generation also induces the reader to go beyond the text. Generating questions forces the reader to think about what is already known and what needs to be learned from the text.

Think-Aloud and Self-Explanation. *Think-aloud strategies* involve the overt, verbal expression of the normally covert mental processes in which readers engage when they are constructing meaning from text (Ericsson & Simon, 1980, 1998). Readers report whatever comes to mind as they read and construct meaning from a text. Think-aloud originally attracted attention among researchers as a potential method to gain insight into the cognitive processes people use to solve problems or read texts, in particular, challenging texts (see Kucan & Beck, 1997; Pressley & Afflerbach, 1995). Subsequently, think-aloud was recognized as a useful method to improve comprehension; that is, there is evidence that when people overtly explain the reading process to themselves (i.e., engage in self-explanation), reading comprehension tends to improve (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; see also Kucan & Beck, 1997, for review of the effects of think-aloud on reading comprehension).

There is some question, however, concerning the extent to which think-aloud, as opposed to self-explanation, helps to improve comprehension. Think-aloud alone may not be particularly effective in influencing the comprehension process. Nonetheless, it is a first step toward the process of self-explanation. Although few readers naturally self-explain well, research has shown that the self-explanation can be successfully taught (Chi, De Leeuw, Chiu, & LaVancher, 1994; McNamara, 2004; chap. 16, this volume). There are several possible reasons why self-explaining text improves reading comprehension (see chap. 16, this volume). First, self-explanation tends to engage readers into a mode of active processing. Second, self-explanation helps readers to detect problems in comprehension by externalizing internal thought processes. As a result, self-explanation can lead readers to repair comprehension problems.

Self-explanation also helps readers to engage in *elaboration*, which involves relating the text content to what one already knows. This process plays at least two roles in facilitating deep processing of the text: (a) filling the conceptual gaps in the text and (b) integrating the text content into pre-existing knowledge structures. Pressley and Afflerbach's (1995) review of think-aloud protocol analysis of skilled reading behavior indicated that elaboration can occur in a number of different ways. For example, a reader can explain an event or fact mentioned in the text by adding relevant details and attributes that he or she knows on the basis of prior knowledge. Or, a reader can think about examples of the concepts presented in the text based on personal experiences or information acquired from other texts.

Visualization. *Visualization* or *imagery* is a useful way of going beyond the text, because visualizing induces the reader to draw on prior knowledge and thus ground ideas discussed in the text. Also, visual memory is sometimes better than verbal memory (e.g., Shepard, 1967). Visualization is a particularly useful strategy for younger readers and for narrative texts whereupon images of familiar situations can easily be evoked. Research in the area of imagery has been promising: Participants who are asked to visualize text contents remembered and comprehended more than control students (Center, Freeman, Robertson, & Outhred, 1999; chap. 9, this volume; McCallum & Moore, 1999; Oakhill & Patel, 1991).

Role playing and manipulating objects to re-enact the text may be first steps toward the processes of visualization and imagery (e.g., chaps. 9 and 12, this volume). Role playing or acting out parts of narratives or plays, for example, grounds the textual experience into something that is tangible and exemplified in experience. Similarly, comprehension through action is central to Glenberg's theory of embodied cognition (chap. 9, this volume). The central claim of the embodiment assumption is that symbols and language are grounded in action (Glenberg, Havas, Becker, & Rinck, 2005). In other words, comprehension and memory should improve when readers are able to experience or act out the information reported in the text; grounding the text into real experience helps readers form a more complete situation model.

Using Outside Sources. In a number of reading situations, in particular when reading expository texts, such as science texts with unfamiliar content, readers often encounter incomprehensible material. For example, a sentence may contain several words whose meanings are completely unknown to the readers. Readers can try to guess the meaning of the sentence based on the contextual cues, such as the meaning of the other words or the preceding discourse context. However, ultimately there are circumstances in which readers need to turn to external sources to find the meaning of the words. There are numerous primary and secondary sources for looking up information, such as word meanings, facts, and phenomena. For example, students may hold discussions or ask an expert (e.g., teacher) to find out information. Alternatively, students can search dictionaries for word meanings, or consult encyclopedias for information on various concepts.

Secondary sources, such as making reference to written materials, play a particularly important role in helping students find out new information. One strategy is to use a dictionary or thesaurus to look up new words (Stahl & Fairbanks, 1986). A dictionary provides information about the meaning, pronunciation, and spelling of a word, and a thesaurus contains synonyms for commonly used words. However, dictionaries and thesauruses usually provide information about nontechnical words, and thus the reader should consult a glossary in the book for technical word definitions. Using reference material to look up word meanings is a strategy that can be used fairly early in children's reading development.

Secondary sources are also useful for looking up facts or deepening the reader's understanding of concepts and ideas. Readers may use an encyclopedia or almanac to look up more detailed information about concepts and phenomena. Readers have a range of options about how to access secondary reference material. They may use copies of reference materials (e.g., dictionaries) from the classroom or libraries. The Internet is a widely available tool in contemporary educational establishments that gives the reader access to multiple sources of information. For example, students may access online dictionaries or encyclopedias. Furthermore, the reader may use search engines to find out about particular concepts and ideas contained in a text.

The critical idea is that the reader should not leave comprehension problems unresolved; instead, he or she should use available sources to solve problems to attain the level of comprehension required for the circumstances. By referring to external sources, readers can not only comprehend the text but also learn new information through the searching activities.

Unfortunately, to our knowledge no research has examined the issue of precisely what kind of search techniques help readers better comprehend and learn from text. There are a number of important questions about this strategy, including whether readers should suspend the reading process and engage in the searching process as soon as they encounter the problem, or whether they should wait until they reach the end of relevant section, and whether nontextual materials (e.g., diagrams or pictures) are useful in helping the readers overcome the difficulty. What would be the most effective search strategy? Perhaps the answers to these questions may depend, to some extent, on the specific circumstances of reading. Future research should explore these questions, because the need to refer to external sources is quite frequent in many reading situations, particularly for novice readers faced with challenging materials.

Strategies to Organize, Restructure, and Synthesize the Text Content

- The reader uses graphic organizers and reading guides.
- The reader engages in writing activities such as summarization.

Successful reading comprehension of extended texts involves the formation of a well-organized and coherent mental representation of the overall text content. In effect, such a representation must have a network-like structure in which relations among the concepts, ideas, and other various attributes of the text are represented in a coherent fashion (Kintsch, 1998). However, the formation of such an optimally organized mental representation of the text is difficult, for several reasons. First, many texts do not provide information in an optimally organized fashion (Beck et al., 1991), and therefore reading a text and relying on structural cues provided in the text do not automatically lead to the formation of a well-organized representation (Britton, Gulgoz, & Glynn, 1993).

Second, readers often need to organize information provided in the text in a specific way to prepare for the tasks that they expect to perform based on text content (e.g., answering questions, writing essays, preparing questions). Mentally organizing the textual content in a specific way requires actively imposing an organizational scheme that fulfills the reader's own goals. However, for many readers who are already struggling with the process required to understand individual sentence meanings or local text meaning, reorganizing the textual content at the time of processing the text is overwhelming. As such, it is very likely to result in interference among competing demands. This scenario is highly likely when students need to read and understand instructional texts relating to content domains (e.g., biology, chemistry, social studies), which are designed to provide unfamiliar information to students. Because of these reasons, it is important to emphasize that reading comprehension involves far more than online processing that occurs at the time of actually reading individual sentences. It also requires engaging in tasks after reading, which serve to consolidate the information in the text.

Thus, as discussed by Magliano et al. (chap. 5, this volume), reading comprehension involves processing that occurs after readers finish reading the text. Reading strategies or efforts to improve the understanding of text materials need to continue after the reading process has been completed, because text processing while reading the text alone often does not lead to the formation of a coherent representation of the global text content. Readers need to reprocess the information learned from the text through various postreading activities that facilitate further organization and synthesis of the information. There are certainly numerous activities to facilitate global processing and help students organize or synthesize information learned from texts. Here, we include two categories of activities. These activities include (but are not limited to) (a) rerepresenting text content using a graphic or schematic format (e.g., diagram, schematic picture, graphic organizer, table, reading guide) and (b) summary writing.

Using Graphic Organizers and Reading Guides. Whether a text is narrative (e.g., story) or expository (e.g., science text), it presents information to readers in a linear fashion. This follows from the fact that words are presented sequentially. However, the order of presentation of the ideas and concepts in a text usually does not correspond directly to the underlying meaning of the text. For example, the narrative may begin at the end of the story and then present various events that led to the end. Indeed, many narratives go back and forth in time. Expository, or informational text also generally requires the reader to make connections between ideas in a nonlinear fashion (e.g., hierarchical/taxonomic or parallel relations) for a coherent representation of the content to develop in the readers' mind. Readers with knowledge about text structure appear to capture nonlinear complex relations among the concepts and ideas (e.g., comparison, problem and solution, cause and effect, taxonomic relations) in texts using various cues such as connectives (Meyer, Brandt, & Bluth, 1980), headings (Lorch & Lorch, 1996), and macro statements presented at the beginning of the sentences while reading the text (Meyer & Poon, 2001; chaps. 8 and 14, this volume). However, for readers with a poor knowledge of text structure, paying attention to and continuously keeping track of the structure of the text while reading is a demanding task. Two types of strategies can help readers to schematically organize the text in a meaningful way: (a) graphic strategies and (b) reading guides.

Graphic representations of the text contents can capture the overall macro structure and relations between important concepts and ideas in the text. The advantage of graphic representation is its ability to represent relatively complex nonlinear relations among the concepts in the text, and rerepresenting the text contents using a graphic format will help readers organize ideas and concepts in the text into more meaningful ways. Graphic organizers also externalize the ideas in the text.

Generally speaking, there are three types of strategies that help readers organize text contents using graphic format: (a) graphic organizers (e.g., Barron & Shwartz, 1984; Griffin, Mallone, & Kamennui, 1995), (b) knowledge maps (Chmielewski & Dansereau, 1998; Dansereau & Newbern, 1997), and (c) concept maps (Novak, 1990; Romance & Vitale, 2001; chap. 4, this volume). We do not describe the differences between these specific strategies in detail here; suffice it to say that they are similar in underlying principles. The creation of a graphic representation converts linear textual statements into nonlinear graphic representation with some sort of tree structure that captures the overall relations between the ideas in the text.

Empirical studies support the benefit of graphic strategies in helping readers' memorization and comprehension of descriptive text contents (e.g., Alverman, 1981), concept acquisition (Robinson, Katayama, Dubois, & DeVaney, 1998), science concept learning (Novak & Musonda, 1991; Schmid & Telaro, 1991; Vitale & Romance, this volume), and learning from texts (Hauser, Nuckles, & Renkl, 2006). Overall, the evidence indicates that graphic strategies help readers comprehend and learn from a variety of texts.

The other strategy known to help readers organize text contents more globally is the use of reading guides. There are different kinds of reading guides, such as dialectical journals; *SOAPSTone*, an acronym used for analyzing text for point of view in terms of Subject, Occasion, Audience, Purpose, Speaker, Tone, Organization, Narrative Style, and Evidence; and *TP-CASTT* (Title, Paraphrase, Connotation, Attitude, Shifts, Title, and Theme), an acronym used specifically for analyzing poems. *Dialectical journal* is another name for a double-entry journal. Readers write down sentences, paragraphs, or any section of a text in one column (usually the left side) and their own thoughts, opinions, and questions in a second column. Journal writing helps the reader reprocess and organize text contents.

Unlike SOAPstone, TPCASTT represents the type and sequence of analyses in which readers should engage to gain a deeper understanding of a poem. The benefit of using guides such as SOAPSTone and TPCASTT is that they provide a schematic guideline about the attributes to which readers should attend when reading text that thus helps readers to form a more coherent or organized mental representation of text contents. As a consequence, guides provide readers with some specific goals that scaffold their processing effort. To the extent that many readers fail to have specific goals that lead to successful comprehension of texts, guides help readers to process text analytically by providing immediate and specific goals.

Although the schematic organization strategies, such as the use of guides and graphic organizers, are effective, one may need to adjust the specific application depending on the student's grade and reading level. Furthermore, two issues warrant attention with regard to the use of schematic organization strategies. First, the strategies covered here aid readers' comprehension mainly at the macro-structure level. In this sense, these strategies should be considered relatively advanced; if readers struggle to comprehend individual sentences or word meanings, then the strategy is unlikely to provide efficient help. Thus, the application and/or training of graphic strategies and use of guides should take place with a text in which students can develop sufficient sentence-level understandings. Second, students should ultimately be able to generate their own reading guides and graphic representations that are tailored according to specific reading goals of individual readers. However, having struggling readers generate concept maps of a science text from scratch is not realistic. Thus, training should follow the path of gradual withdrawal of scaffolding, through which students learn text contents with the support of graphic representations (e.g., concept maps) constructed by experts and then move toward constructing their own graphic representations (Chang, Sung, & Cheng, 2002; Hauser et al., 2006).

Summary Writing. Summary writing contributes to deeper comprehension of texts in several different ways. First, similar to paraphrasing, which we described earlier in this chapter, summary writing involves restating text content in the reader's own words and expressions, increasing the chance that the information explicitly stated in the text becomes integrated with the reader's pre-existing knowledge (Wade-Stein & Kintsch, 2004). Second, summarizing text content forces readers (and teachers) to be sensitive to whether reading materials are understood and where there are comprehension difficulties. Identifying comprehension difficulties can in turn guide subsequent remedial actions to improve comprehension of the materials (e.g., rereading specific sections of the text that cannot be recalled from memory). Third, summary writing helps readers organize text contents at the macro level and discern which information is important and which is not. Finally, summary writing serves as a foundation for more complex, creative processes (e.g., critical essay, report, research paper writing) that people are often required to perform in academic and professional settings.

Studies indicate that the quality of summary writing mirrors the reader's comprehension of the text. For example, the summary writing of struggling comprehenders often shows a lack of consistency associated with the progression of a topic (e.g., McCutchen & Perfetti, 1982) and a violation of the hierarchical organization of the text (e.g., Meyer et al., 1980). In addition, evidence indicates that summarization results in better comprehension (e.g., Bean & Steenwyk, 1984; Casazza, 1992; Moore, 1995) and better retention of text materials (e.g., Rinehart, Stahl, & Erickson, 1986). These findings suggest that practicing summary writing is beneficial for improving reading comprehension ability. Caccamise et al. (chap. 15, this volume) showed that the interactive practice of summary writing through Summary Street, a computerized summary writing trainer, resulted not only in the production of better summary writing but also in better comprehension of texts. This suggests that guided practice of summary writing exerts a positive effect on text processing. Thus, overall, the literature indicates that both summary writing and providing practice of summary writing contributes to improvement in reading comprehension.

One issue that remains unclear is for whom and at what stage of development summary writing practice is most beneficial. As fuchs and Fuchs outline in chapter 7 of this volume, children may begin to produce summaries at the elementary level. The authors explain how a peer-mediated task called *Peer-Assisted Learning Strategies* supports the production of accurate summaries. The Peer-Assisted Learning Strategies technique requires a child reader and tutor (more experienced child) to jointly construct paragraph summaries. The tutor helps the reader by pointing out errors the reader has made after reading aloud a paragraph and summarizing who or what the paragraph was mostly about and the most important thing that happened. However, the issue regarding whom and when to teach summary writing is complicated by the possibility that the benefit of the training may depend on the nature of training, such as the type of text used and the type of feedback provided. For example, the findings of Caccamise et al. (chap. 15, this volume) suggest that the benefit of summary writing is larger for readers of low and intermediate ability reading challenging texts. Nonetheless, depending on the nature of the training practice, summary writing practice may benefit more advanced readers by helping them acquire skills to integrate text-based information with prior knowledge in a more extensive manner. Readers can learn to construct more advanced forms of summarizing, such as comparative or integrated summaries that go beyond the constraints of the source text. Although these are certainly steps in the right direction, more research is needed to fully resolve the parameters influencing the benefits of summarization.

CONCLUSION

In this final chapter, we have introduced a framework to organize various strategies to improve comprehension. We are aware that we have ignored potentially important strategies. However, our goal was not to include all strategies but rather to include a majority of the ones believed to be effective and, moreover, to provide an organizing schema for thinking about reading strategies.

Each of these strategies is going to be more or less effective depending on the reader, the text, and the circumstance. An important ingredient in becoming a skilled reader is not just learning to use strategies but learning when to use which strategies. More research is needed so that we can better understand when certain strategies will be more or less effective. To some extent, many of the chapters in this volume contribute to this understanding. However, we look forward to more work to resolve the question of what readers can and should do, and when.

ACKNOWLEDGMENTS

We are grateful to Phil McCarthy, Margie Petrowski, and Nick Duran, who provided comments on an earlier version of this chapter. The preparation of this chapter was partially funded by the Institute of Education Sciences Reading Program (R305G040046). The framework presented in this chapter was prepared in collaboration with the College Board as part of its development of the English Language Arts College Board Standards for College Success™, published in 2006 by the College Board. As such, the framework was developed with the help of many individuals who served on the College Board English Language Arts Standards Advisory Committee; a complete list of the contributors is presented in the standards document. We are particularly grateful to those who served on the Reading Standards Advisory Committee, including Charlie Peters; Judson Odell; Nina Wooldridge; Cathy Roller; Gary Cowan; and particularly Arthur VanderVeen, who led the committee. Readers interested in how the strategies might progress developmentally should refer to the standards document, which presents six levels of expected performance through middle school and high school. This chapter, in contrast to the standards, describes the theoretical and empirical foundations of the framework.

REFERENCES

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge. *Review of Educational Research*, *58*, 375–404.

Alverman, D. E. (1981). The compensatory effect of graphic organizers on descriptive text. *Journal of Educational Research*, *75*, 44–48.

Anderson, R. C., & Pearson, P. (1984). Research methodology on the use of verbal reports in reading research. *Journal of Reading Behavior*, *16*, 307–322.

Anderson, T. H., & Armbruster, B. B. (1986). *The value of taking notes during lectures* (Technical Report No. 374). Cambridge, MA: University of Illinois at Urbana-Champaign, Bolt, Beranek and Newman Center for the Study of Reading.

Anderson, V., & Hidi, S. (1988–1989). Teaching students to summarize. *Educational Leadership*, *46*, 26–28.

Baker, L. (1985). Differences in the standards used by college students to evaluate their comprehension of expository prose. *Reading Research Quarterly*, *20*, 297–313.

Barron, R. F., & Shwartz, R. M. (1984). Traditional post organizers: A spatial learning strategy. In C. D. Holley & D. F. Dansereau (Eds.), *Spatial learning strategy, techniques, applications, and related issues* (pp. 275–289). New York: Academic.

Bean, T. W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Reading Behavior*, *16*, 297–306.

Beck, I., & McKeown, M. (2001). Inviting students into the pursuit of meaning. *Educational Psychology Review*, *13*, 225–241.

Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman, J. A. (1991). Revising social studies text from a text processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly*, *26*, 251–276.

Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction*, *2*, 131–156.

Bradtmueller, W. (1983). To question or not to question: That seems to be the question. *Journal of the Association for the Study of Perception*, *18*, 23–28.

Britton, B. K., Gulgoz, S., & Glynn, S. (1993). Impact of good and poor writing on learners: Research and theory. In B. K. Britton, A. Woodward, & M. R. Binkley (Eds.), *Learning from textbooks: Theory and practice* (pp. 1–46). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brown, A. (1982). Learning how to learn from reading. In J. A. Langer & M. T. Smith-Burke (Eds.), *Reader meets author: Bridging the gap* (pp. 26–54). Newark, DE: International Reading Association.

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference making ability and their relation to knowledge. *Memory & Cognition*, *29*, 850–859.

Casazza, M. E. (1992). Teaching summary writing to enhance comprehension. *Reading Today*, *9*, 28.

Casteel, C. (1989). Effects of chunked text-material on reading comprehension of high and low ability readers. *Reading Improvement*, *27*, 269–275.

Cavanaugh, J. C., & Perlmutter, M. (1982). Metamemory: A critical examination. *Child Development*, *53*, 11–28.

Center, Y., Freeman, L., Robertson, G., & Outhred, L. (1999). The effect of visual imagery training on the reading and listening comprehension of low listening comprehenders in Year 2. *Journal of Research in Reading*, *22*, 241–256.

Chang, K., Sung, Y., & Chen, I. (2002). The effect of concept mapping to enhance text comprehension and summarization. *Journal of Experimental Education*, *71*, 5–23.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145–182.

Chi, M. T. H., De Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self explanations improves understanding. *Cognitive Science*, *18*, 439–477.

Chmielewski, T., & Dansereau, D. F. (1998). Enhancing the recall of text: Knowledge mapping promotes implicit transfer. *Journal of Educational Psychology*, *90*, 407–413.

Clark, F. L., Deshler, D. D., Schumaker, J. B., Alley, G. R., & Warmer, M. M. (1984). Visual imagery and self-questioning: Strategies to improve comprehension of written material. *Journal of Learning Disabilities*, *17*, 145–149.

Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, *25*, 1–53.

Dansereau, D. F., & Newbern, D. (1997). Using knowledge map to enhance teaching. In W. E. Campbell & K. A. Smith (Eds.), *New paradigms for college teaching* (pp. 125–147). Edina, MN: Interaction Book.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87,* 215–251.

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, & Activity*, *5*, 178–186.

Everson, H., & Tobias, S. (1998). The ability to estimate knowledge and performance in college: A metacognitive analysis. *Instructional Science*, *26*, 65–79.

Glenberg, A. M., Havas, D., Becker, R., & Rinck, M. (2005). Grounding language in bodily states: The case for emotion. In R. Zwaan & D. Pecher (Eds.), The grounding of cognition: The role of perception and action in memory, language, and thinking (pp. 115–128). Cambridge, England: Cambridge University Press.

Glenberg, A., Wilkinson, A., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, *10*, 597–602.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.

Griffin, C. C., Malone, L. D., & Kameenui, E. J. (1995). Effects of graphic organizer instruction on fifth grade students. *Journal of Educational Research*, *75*, 275–279.

Haller, E., Child, D., & Walberg, H. J. (1988). Can comprehension be taught? A quantitative synthesis. *Educational Researcher*, *17*, 5–8.

Hannon, B., & Daneman, M. (1988). Facilitating knowledge-based inferences in less-skilled readers. *Contemporary Educational Psychology*, *23*, 149–172.

Hauser, S., Nuckles, M., & Renkl, A. (2006). Supporting concept mapping for learning from text. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *Proceedings of the Seventh International Conference of the Learning Sciences* (pp. 243–249). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Herrell, A., & Jordan, M. (2002). *Active learning strategies for improving reading comprehension*. Upper Saddle River, NJ: Merrill Prentice Hall.

Inagaki, K., & Hatano, G. (1974). Correlates of induced question-asking behavior. *Japanese Psychological Research*, *16*, 50–57.

Jonassen, D. H. (2000). *Computers as mindtools for schools: Engaging critical thinking* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Kamberelis, G., & Bovino, T. D. (1999). Cultural artifacts as scaffolds for genre development. *Reading Research Quarterly*, *34*, 138–170.

Keenan, S. (1984). Effects of chunking and line length on reading efficiency. *Visible Language*, *18*, 61–80.

King, A. (1989). Effects of self-questioning training on college students' comprehension of lectures. *Contemporary Educational Psychology*, *14*, 366–381.

King, A. (1994). Autonomy and question asking: The role of personal control in guided student-generated questioning. *Learning and Individual Differences*, *6*, 163–185.

King, A., & Rosenshine B. (1993). Effects of guided cooperative questioning on children's knowledge construction. *Journal of Experimental Education*, *61*, 127–148.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.

Kozminsky, E. (1992, April). *Go for GOs—Cognitive processes underlying the use of graphic organizers*. Paper presented at the meeting of the American Educational Research Association, San Francisco.

Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, *67*, 271–299.

Kurtz, B. E. (1991). Cognitive and metacognitive aspects of text processing. In G. Denhiere & J. P. Rossi (Eds.), *Text and text processing* (pp. 77–103). New York: North-Holland Elsevier.

Langer, J. A. (1986). Relations between levels on prior knowledge and the organization of recall. In M. L. Kamil & A. J. Moe (Eds.), *Perspectives in reading research and instruction* (pp. 28–33). Washington, DC: National Reading Conference.

Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1456–1470.

Lorch, R. F., Jr., & Lorch, E. P. (1996). Effects of headings on text recall and summarization. *Contemporary Educational Psychology*, *21*, 261–278.

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure. *Cognition and Instruction*, *3*, 251–283.

Maqsud, M. (1997). Effects of metacognitive skills and nonverbal ability on academic achievement of high school pupils. *Educational Psychology*, *17*, 387–397.

Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. *Child Development*, *48*, 986–992.

McCallum, R., & Moore, S. (1999). Not all imagery is created equal: The role of imagery in the comprehension of main ideas in exposition. *Reading Psychology*, *20*, 21–60.

McCutchen, D., & Perfetti, C. (1982). Coherence and connectedness in the development of discourse production. *Text*, *2*, 113–139.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440–466.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*, 1–30.

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–287.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1–43.

Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly*, *16*, 72–103.

Meyer, B. F., & Poon, L. W. (2001). Effects of structure strategy training and signaling on recall of text. *Journal of Educational Psychology*, *93*, 141–159.

Moore, S. R. (1995). Focus on research questions for research for research into reading–writing relationships and text structure knowledge. *Language Arts*, *72*, 598–606.

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching*, *27*, 939–949.

Novak, J. D., & Musonda, D. (1991). A twelve year longitudinal study of science concept learning. *American Educational Research Journal*, *28*, 117–153.

Novak, J. D. (2004). *The theory underlying concept maps and how to construct them*. Retrieved April 8, 2004, from http://cmap.coginst.uwf.edu/info/printer.html

Oakhill, J. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, *54*, 31–39.

Oakhill, J. V., & Patel, S. (1991). Can imagery training help children who have comprehension problems? *Journal of Research in Reading*, *14*, 106–115.

Oakhill, J., & Yuill, N. (1996). Higher order factors in comprehension disability: Processes and remediation. In C. Cornaldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 69–92). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

O'Reilly, T. P., & McNamara, D. S. (in press). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high stakes" measures of high school students' science achievement. *American Educational Research Journal*.

O'Reilly, T., Symons, S., & MacLatchy-Gaudet, H. (1998). A comparison of self-explanation and elaborative interrogation. *Contemporary Educational Psychology*, *23*, 434–445.

Penno, J. F., Wilkinson, A. G., & Moore, D. S. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology*, *94*, 23–33.

Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Volume III* (pp. 545–561). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Pressley, M. (2002). Metacognition and self-regulated comprehension. In A. E. Farstrup & S. Samuels (Eds.), *What research has to say about reading instruction* (pp. 291–309). Newark, DE: International Reading Association.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Pressley, M., & Ghatala, E. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist*, *25*, 19–33.

Richardson, J. S., & Morgan, R. F. (2000). *Reading to learn in the content areas*. Belmont, CA: Wadsworth.

Rinehart, S. D., Stahl, S. A., & Erickson, L. G. (1986). Some effects of summarization training on reading and studying. *Reading Research Quarterly*, *21*, 422–437.

Robinson, D. H., Katayama, A. D., Dubois, N. F., & DeVaney, T. (1998). Interactive effects of graphic organizers and delayed review on concept acquisition. *Journal of Experimental Education*, *67*, 17–31.

Romance, N. R., & Vitale, M. R. (2001). Implementing an in-depth expanded science model in elementary schools: Multi-year findings, research issues, and policy implications. *International Journal of Science Education*, *23*, 373–404.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, *64*, 479–530.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, *66*, 181–221.

Rowe, M., Ozuru, Y., & McNamara, D. S. (2006). An analysis of s standardized reading ability test: What do questions actually measure? In S. A. Barab, K. E. Kay, & D. T. Hickey (Eds.), *Proceedings of the Seventh International Conference of the Learning Sciences* (pp. 627–633). Mahway, NJ: Lawrence Erlbaum Associates.

Saricoban, A. (2002). Reading strategies of successful readers through the three phase approach. *Reading Matrix*, *2*, 1–16.

Schmid, R. F., & Telaro, G. (1991). Concept mapping as an instructional strategy for high school biology. *Journal of Educational Research*, *84*, 78–85.

Schmitt, M. (1990). A questionnaire to measure children's awareness of strategic reading process. *The Reading Teacher*, *49*, 454–461.

Schraw, G., & Dennison, R. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, *19*, 460–475.

Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal*, *41*, 159–189.

Shepard, R. (1967). Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156–163.

Spires, H. A., Gallini, J. K., & Riggsbee, J. (1992). Effects of schema-based and text structure-based cues on prose comprehension in fourth graders. *Journal of Experimental Education*, *60*, 307–320.

Stahl, S., & Fairbanks, M. (1986). The effects of vocabulary instruction: A model based on meta-analysis. *Review of Educational Research*, *56*, 72–110.

Sternberg, R., & Powell, J. (1983). Comprehending verbal comprehension. *American Psychologist*, *38*, 878–893.

Thiede, K. W., Anderson, M. C., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*, 66–73.

Todaro, S., Magliano, J. P., Millis, K. K., Kurby, C. A., & McNamara, D. S. (2006). *Understanding factors that influence the content and form of verbal protocols: The roles of the reader and text*. Submitted.

van den Broek, P., Lorch, R. F., Jr., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*, 1081–1087.

van Kraayenoord, C., & Schneider, W. (1999). Reading achievement, metacognition, reading self-concept and interest: A study of German students in grades 3 and 4. *European Journal of Psychology of Education*, *14*, 305–324.

Vellutino, F. R. (2003). Individual differences as sources of variability in reading comprehension in elementary school children. In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 51–81). New York: Guilford.

Wade, S. E., Trathen, W., & Schraw, G. (1990). An analysis of spontaneous study strategies. *Reading Research Quarterly*, *25*, 147–166.

Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*, *22*, 333–362.

Williams, J. P., Brown, L. G., Silverstein, A. S., & deCani, J. S. (1994). An instructional program for adolescents with learning disabilities in the comprehension of narrative themes. *Learning Disability Quarterly*, *17*, 205–221.

Williams, J. P., Hall, K. M., & Lauer, K. D. (2004). Teaching expository text structure to young at-risk learners: Building the basics of comprehension instruction. *Exceptionality*, *12*, 129–144.

Williams, J. P., Lauer, K. D., Hall, K. M., Lord, K. M., Gugga, S. S., Bak, S. J., et al. (2002). Teaching elementary school students to identify story themes. *Journal of Educational Psychology*, *94*, 235–248.

Yuill, N., & Joscelyne, T. (1988). Effect of organizational cues and strategies on good and poor comprehenders' story understanding. *Journal of Experimental Psychology*, *80*, 152–158.

Author Index

C

D

E

F

G

N

O

Z

Subject Index

T

"What an exciting discovery, to find a book containing so many outstanding chapters on reading comprehension, many of which were written by the world's most famous scholars on this topic. For those who want to find out how the mind works as the reader engages in the most mysterious task of making sense out of what is printed on a page, and for those who want to find ways to increase the reading comprehension of their students, this book is essential reading."

—Jay Samuels, National Reading Panel

"This volume represents the state of the art in connecting theories of readers' cognitive processes with practical interventions for improving their comprehension. In a field of study that is notoriously complex and hard to synthesize, *Reading Comprehension Strategies: Theories, Interventions, and Technologies* is a useful guidebook to what is known, how we might best conceive of the problems, and what kinds of new approaches are finding empirical support. The serious exploration of how new technologies can support readers of different sorts in becoming better readers will be particularly intriguing for anyone concerned with the future of literacy education."

—Michael Schober, New School for Social Research

"There is no doubt that reading is fundamental, and basic literacy skills are critical for learning to read, but the ultimate goal of reading is comprehension. This collection discusses how we may move readers past 'literal' comprehension toward conceptual understanding of content-area knowledge. The chapters focus on strategy instruction and comprehension assessment and serve as excellent contributions to our knowledge of how we may best promote 'reading to learn.'"

—Jennifer Wiley, University of Illinois at Chicago

"*Reading Comprehension Strategies: Theories, Interventions, and Technologies* presents a readable account of the current state of comprehension research. The authors represent some of the most productive researchers in the field, each discussing two critical concerns that are traditionally considered separately: (1) They outline their specific investigations into the mechanisms that underlie comprehension, and (2) they assess ways in which the operation of these mechanisms might indicate useful remediative strategies and technologies to address comprehension difficulties. The resulting volume is an excellent resource for individuals in the diverse fields of education, cognition, and instructional design."

—David N. Rapp, Northwestern University

ISBN 0-8058-5967-5
T
90000
9 780805 859676

Lawrence Erlbaum Associates
Taylor & Francis Group
www.psypress.com

an **informa** business