International Electronic Journal of ELEMENTARY EDUCATION

www.iejee.com
October 2011 / Volume: 4 Number: 1

Special Issue
Reading Comprehension

International Electronic Journal of Elementary Education
Vol. 4, Issue 1, October, 2011.

IEJEE
International Electronic Journal of Elementary Education
International Electronic Journal of Elementary Education is an international educational periodical that is online published three times in a year. This journal provides immediate open access to its content on the principle that making research freely available to the public supports a greater global exchange of knowledge.

ISSN: 1307-9298

Abstracting and Indexing:

SciVerse® Scopus, ProQuest, DOAJ, Education Research Complete™, Google Scholar, EDNA, Index Copernicus, Norwegian Social Science Data Services, NewJour, Open J-Gate.

All responsibility for statements made or opinions expressed in articles lies with the author.

ISSN:1307-9298

www.iejee.com

International Electronic Journal of Elementary Education, 2011, Vol. 4, Issue 1

Contents

Dear Readers,

This is the third special issue of *International Electronic Journal of Elementary Education (IEJEE)*. The main topic is *Reading Comprehension,* one of the important skills that the students in the schools of any knowledge society must command. To be able to do that the students at every grade in the school must be provided educational conditions. This necessitates teachers who pursue their job in accordance with research based knowledge and pedagogical ideas. It's the scientific circles' duty to deliver a such knowledge and ideas as a condition for the teachers' professional development.

Our conception of reading in general and reading comprehension in particular has improved by the enormous reading research activities during the last decades. Therefore it was a right decision by IEJEE to focus on this important topic.

Under the editorialship of Dr. Karen M. Zabrucky, Georgia State University, USA, thirty two academicians with affiliation to seventeen different universities, from six countries and three continents, have addressed several aspects of reading comprehension. I do not have any doubt that this special issue of IEJEE will be an important contribution to the field of reading research international.

Dr. Karen M. ZABRUCKY has done a great job. I want to express my deep gratefulness to her for materializing this special issue on a such important topic. I also want to thank to my closest colleagues **Dr. Turan TEMUR**, Dumlupinar University, Turkey and **Dr. Gökhan ÖZSOY**, University of Aksaray, Turkey, for their enormous works that they have done as the executive members of the editorial board of IEJEE.

Prof. Dr. Kamil Özerk, Editor -in- Chief
University of Oslo
NORWAY

International Electronic Journal of Elementary Education, 2011, 4(1), 1-4.

Introduction to the Special Issue Research on Reading Comprehension: Past, Present, and Future

Karen M. ZABRUCKY
Georgia State University, USA

It is my great pleasure to introduce the International Electronic Journal of Elementary Education (IEJEE) special issue on reading comprehension, the third special issue since the journal was founded. I believe that readers will find the present group of articles to be highly informative, timely, and interesting. The special issue includes a broad range of articles on reading comprehension and a mixture of theoretical, review, and empirical articles that provide a historical context for research in reading comprehension, allow an understanding of current methodologies and instructional techniques used to study comprehension, address effective comprehension instruction and applications, and allow a greater understanding of the most promising avenues for future research.

When asked if I would be interested in being the guest editor of the IEJEE special issue on reading comprehension, I experienced little hesitation and much excitement (as well as some nostalgia). Being an undergraduate student in the early 1970s I can point to the class at Kent State University that led to my life-long interest in reading comprehension. D. James Dooling, who happened to be working on text comprehension at the time, taught the course. His enthusiasm was contagious and students in his class were soon doing their best to understand the rather strange but wondrous passages he brought to class. What led to our understanding of the passages? How could a mere title (e.g., Christopher Columbus) make one version of a passage easy to understand while the absence of the title produce a seemingly incomprehensible text? Why were individuals able to recall texts better than the word lists that other psychologists seemed so fond of studying for so long? Although I enjoyed all of my research experiences as an undergraduate psychology student, including work on memory processes with David C. Riccio (although I confess I was not overly fond of working with rats), my interest in reading comprehension won out and my sense of wonder about how comprehension occurs never left.

It is with a great deal of enthusiasm that I introduce this special issue. Although papers in the special issue are diverse, several important themes emerge. One theme involves what comprehension is; what is the essential essence of comprehension? Other themes are whether comprehension is similar in reading versus other contexts in which it can occur (or not), and whether we can use our understanding of what comprehension is to develop effective instructional techniques.

Arthur M. Glenberg's research on *embodied comprehension* addresses all of these themes in an interesting and creative manner (and thanks again Art for getting your paper to me in July!) and describes an effective technique for teaching comprehension called *Moved By Reading*. As the name of his instructional technique implies, comprehension is fostered through the pairing of textual information and action. Children read stories that relate to particular scenarios and objects and either have the objects and characters described in the scenario available to them physically or on a computer screen. When children read each sentence of a story out loud, they use physical manipulation to act out the information with objects or computer images. *Physical manipulation*, as well as a transfer procedure called *imagined manipulation*, results in dramatic increases in comprehension. In his paper, Glenberg describes various contexts in which the *Moved by Reading* intervention was used quite successfully.

Caitlin McMunn Dooley has contributed a review of the literature on *emergent comprehension*, a topic that is gaining more attention in the field. In her article she examines how best to think of comprehension in very young children and argues that conventional models of comprehension do not truly capture the experiences of young children. She proposes additional components, such as relational and symbolic interactions, that must be taken into account when examining emerging comprehension and provides a review of the literature in this area during the last decade. In related work, Paul van den Broek, Panayiota Kendeou, Sandra Lousberg and Gootje Visser discuss their empirical work on comprehension in very young children, Van den Broek and colleagues examined question-asking interventions with early readers as well as toddlers and report on the results of their findings.

Several authors in the special issue addressed the concept of text structure and the important role it plays in enhancing or hindering comprehension. Children are exposed to certain types of text structures from very young ages, such as those found in narratives or stories, and develop well-formed narrative schemas that allow them to encode, store, and retrieve passage information relatively easily. In the case of other types of text structures, such as those occurring in expository or explanatory texts, this is simply not true. Several investigators in the special issue argue that children get very little exposure to expository or informational texts in their early years, little training on the structure of expository texts or the strategies needed to comprehend them, and are only exposed to such texts as they progress in schools. Such problems directly address two issues that emerged in authors' writings. First, several authors commented on the necessity for children to be exposed more frequently to expository texts at a younger age. Additionally, several authors commented on the necessity of students understanding the structure (or, more accurately, structures) of expository texts.

In an article on the importance of knowledge of a variety of structures used in expository texts, Bonnie J. F. Meyer and Melissa N. Ray provide a rich and thorough historical review as well as an update of *strategy structure research* and *strategy structure interventions*. The authors begin their paper by examining early and basic research on structure strategies, and then review both early as well as recent effective intervention studies based on strategy structure research. The authors also provide a discussion of a variety of expository structures that individuals of all ages must become familiar with in order to allow for easier text comprehension and better text recall. In a second paper, Ray and Meyer discuss the role of selected reader characteristics (comprehension skill, age, and prior knowledge) and text characteristics (the hierarchical organization of higher and lower order propositions as well as three types of expository text structure) when examining the literature on individual differences in younger and older children's knowledge and use of expository text structures.

In several papers authors discuss expository texts that convey science content, as such texts may have factors over and above other expository texts that hinder comprehension. Danielle S. McNamara, Yashuhiro Ozuru, and Randy G. Floyd report on a study in which they examined the role of text genre (narratives and expository texts containing science content), text cohesion (high and low) and selected individual difference variables (children's dcoding skills and world knowledge) on reading comprehension, to help explain what has been referred to as the *fourth grade slump*. In a further exploration of the characteristics of science texts, Dianna J. Arya, Elfrieda H. Hiebert and P. David Pearson discuss how lexical as well as syntactic complexity affect children's comprehension. Brandi E. Johnson-Lee and Karen M. Zabrucky review a variety of intervention strategies that have been used successfully with older children (middle-school and high-school students) to help students better understand science texts.

In a unique longitudinal study, Linda Baker, Mariam Jean Dreher, Angela Katenkamp, Lisa Carter Beall, Anitia Voelker, Adia Garrett, Heather R. Schugar, and Maria Finger-Elam report on *The Reading, Engaging, and Learning project (REAL)*. The investigators began working with students when the children were in the second grade and followed children until the end of the fourth grade. The primary purpose of the investigation was to determine whether the infusion of informational books within classroom libraries as well as specific types of "reading for learning" instruction by teachers would enhance children's reading comprehension, achievement, and engagement. The study is informative not only with regard to specific research findings but also with regard to lessons in unpredictable and unforeseen events when conducting ecologically valid research in real world contexts.

Thus, many of the researchers who contributed to the special issue were concerned with issues of how much exposure to expository/informational texts children receive in their classrooms at an early age as well as how much exposure children have to expository text structure and strategies to help "decode" the structure so that they may better understand textual information. There was a particular focus on expository texts within a science domain, as such texts appear to be particularly problematic for children. Baker et al.'s work indicates that there are important differences in teachers' frequency and manner of use of expository texts, even when such texts have been provided to classroom libraries. These findings reveal the complexity of issues that researchers must take into account when trying to increase younger children's exposure to expository or informational texts.

Some of the researchers in the special issue explored the important role of *working memory* in reading comprehension. Jane Oakhill, Nicola Yuill, and Alan Garnham conducted a study to determine if the relationship between working memory and children's reading comprehension was due to general working memory skills or, rather, modality-specific working memory. They examined children's working memory abilities in three different modalities and studied the relationship of children's skills on those measures of reading comprehension as well as reading accuracy. Hye K. Pae and Rose A. Sevcik explored the role of working memory in children's comprehension and reading fluency. They too used multiple indicators of working memory ability and examined the role of these skills in both first language and second language learners (English-Korean bilinguals in the United States and Korean-English bilinguals in Korea).

In somewhat of a departure from several of the papers in the special issue, Oddny Judith Solheim and Per Henning Uppstad report on a study in which they used an *eye-tracking methdology* in a novel manner. Solheim and Uppstadd discuss methodological issues related to the question of how best to validate inferences on the basis of comprehension test scores and suggest that validation requires different types of converging evidence. The authors collected eye-tracking data during children's first reading of a text as well as when children read the text and answered questions about it. The relations between these data and comprehension were explored. In Yousif Alshumaimeri's research we move from examining a particular type of methodology to investigate inferences to examining Saudi EFL students' preferences for different reading methods (oral, silent, and sub-vocalizing). Alshumaimeri examined whether Saudi students had preferences for reading mode, reasons for preferences if they existed, and effects of reading mode on comprehension.

Finally, Danielle S. McNamara and Panayiota Kendeou identified five critical findings in research on reading comprehension and discussed applications of these findings for educational practice. The research findings and applications the authors discussed were in the general areas of the uniqueness of decoding and comprehension skills early in children's lives, the relations between processes and products in comprehension research and instruction, the importance of the development and fostering of inference skills, the multi-faceted nature of comprehension, and the need to be aware of limitations of standardized tests of reading comprehension. The McNamara and Kendeau paper should be accessible and informative to educators in the field.

I would like to express my gratitude to many individuals for their contributions to the special issue. First, many thanks to the authors for not only contributing such excellent papers but contributing them in the time frame necessary to develop the special issue. Thanks also to several gracious and generous colleagues who reviewed for the special issue in a much shorter than usual time period and who produced very helpful reviews. The names of reviewers appear elsewhere in the issue. And, finally, I would like to thank the editorial team at IEJEE for giving me the opportunity to develop the special issue, which I enjoyed more than I can say. Many thanks to Kamil Ozerk for his kind and comforting words. They served to increase my self-efficacy for the new role I found myself in as well as made me smile. And I am deeply indebted to Turan Temur for his guidance and support. I am not sure that a week went by without Turan receiving an email from me (perhaps a paper attached, perhaps a question, perhaps a request to do something, perhaps a request to undo something). He handled each of my e-mails, questions and requests with kindness, patience and good humour.

So on to the special issue and let the reading comprehension begin. I hope that readers will enjoy the papers in the special issue and learn from them as much as I did.

International Electronic Journal of Elementary Education, 2011, 4(1), 5-18.

How reading comprehension is embodied and why that matters

Arthur M. GLENBERG[*]
Arizona State University and University of Wisconsin-Madison, United States

Abstract

Reading comprehension, much like comprehension of situations and comprehension of oral language, is embodied. In all cases, comprehension is the ability to take effective action on the basis of affordances related to the body, the physical world, and personal goals and cultural norms. In language contexts, action-based comprehension arises from simulating the linguistic content using neural and bodily systems of perception, action, and emotion. Within this framework, a new approach to teaching reading comprehension is described: Teach children how to simulate while reading. The *Moved by Reading* intervention teaches simulation in two stages. In the first stage, physical manipulation, children manipulate toys to simulate the content of what they are reading. After success in physically manipulating the toys, the children are taught to manipulate the toys in imagination. Research demonstrates that both physical and imagined manipulation leads to large gains in memory and comprehension.

Keywords: Reading comprehension, embodiment, Moved by Reading

Introduction

What does it mean to comprehend? Does comprehension differ when understanding situations, oral language, or written texts? Does the nature of comprehension hold any implications for the nature of instruction? These are some of the questions I hope to answer in this essay. In brief, I will propose that comprehension is related to action: Understanding a situation or a text means that the understanding can be used to guide effective action, and that this definition holds whether one is understanding situations, dialogue, or text. Furthermore, because understanding guides literal action, understanding is closely related to bodily abilities. Finally, understanding the embodied nature of reading comprehension suggests an effective technique for teaching reading comprehension skill to even the

[*] Arthur M. Gleberg, Department of Psychology, Arizona State University, Tempe, AZ, 85283-1104, United States. Phone: 480-727-7790. E-mail: glenberg@asu.edu.

ISSN:1307-9298

www.iejee.com

youngest of readers. This technique, called *Moved by Reading*, works by having children physically and cognitively interact with text so that an embodied, action-based understanding of the text is achieved.

Understanding situations and language

Let's begin with a commonsense approach to understanding. When walking down the street, sitting at work, or talking with friends at a ballgame, what might it mean to understand? Understanding is not just a cognitive state. If one were to stop moving and simply think about the situation at hand, others would certainly become alarmed. "Hey buddy! You can't just stand there in the middle of the street!" But it is also the case that in any particular situation, most actions are inappropriate: While at work or school you can't start to dance, shout out loud, pretend to swim, or intensively groom yourself or others. If you did, it would certainly be appropriate for someone to say, "Don't you understand where you are?" Thus, understanding a situation is, at the very least, *revealed* through appropriate action, that is, action that is constrained by the physical situation, bodily capabilities, social and cultural norms, and particular goals. However, I will argue further that the ability to act appropriately does not just reveal understanding but constitutes understanding.

A similar sort of commonsense analysis holds for understanding oral language, as in a dialog. If I say, "Wow, the other day Art freaked out at work: He was trying to swim down the hallway," and you respond with "I really don't like to travel on airplanes," I would be justified to ask, "Hey, did you understand what I said?" Again, understanding is revealed by appropriate action in the situation, here a conversation. Yet, there seems to be a big difference between action that involves large body movements, such as pretending to swim, and action that consists of moving the speech articulators to make noises.

The difference between large body movements and speech is resolved with the concept of simulation (Barsalou, 1999; Barsalou, 2008; Gallese & Lakoff, 2005; Glenberg & Gallese, 2011; Zwaan & Taylor, 2006). According to the simulation theory of language comprehension, language is understood by simulating the situation described by the language, that is, by driving the brain into states that are analogous to the perceptual, action, and emotional states that arise during perception of and acting in the real situation. Thus, understanding, "He was trying to swim down the hallway," utilizes a) visual perceptual systems that are active in seeing a real hallway and seeing real motion, b) motor systems that would be used in swimming, and c) emotional systems that would respond to the odd, upsetting, and perhaps frightening situation of seeing a grown man acting inappropriately.

There is a large and growing literature supporting the claims of simulation theory (for reviews see Glenberg, 2007; Kiefer & Pulvermüller, 2011; Fischer & Zwaan, 2008). Here I will describe just a smattering of that evidence. If simulation theory is correct, then upon hearing verbs such as lick, pick, and kick, there should be somatotopic activation of the motor system that controls the mouth, the hands, and the legs, respectively. That is exactly what was reported by Hauk, Johnsrude, & Pulvermüller (2004) who measured brain activity using functional magnetic resonance imaging (fMRI) while people listened to verbs. Similarly, when reading sentences describing transfer of objects from one person to another, simulation theory predicts greater activity in the neural and muscular systems controlling the hand than when reading about the same objects and people without any mention of action. This differential neural-muscular activity was demonstrated using single-pulse transcranial magnetic stimulation (Glenberg et al., 2008).

There is also evidence supporting the claim from simulation theory that perceptual mechanisms are engaged during language comprehension. For example, when people understand sentences that describe motion, there is greater activity in areas V5/MT (that specializes in the visual perception of motion) than when understanding sentences that describe a visual scene without motion (Rueschemeyer, Glenberg, Kaschak, Mueller, & Friederici, 2010).

Finally, simulation predicts that understanding of language with emotional content involves neural and bodily systems of emotion. Havas, Glenberg, Gutowski, Lucarelli, & Davidson (2010) asked people to read sentences describing happy, sad, and angry events. As predicted by simulation theory, while reading about happy events, there was greater activity in the muscles that control smiling, and when reading about sad and angry events, there was greater activity in the corrugator muscle used in furling the brow. Simulation theory also predicts that if the ability to engage in a simulation is blocked or reduced, then comprehension should suffer. Havas et al. (2010) produced evidence consistent with this prediction using a rather unusual procedure involving cosmetic Botox. Cosmetic Botox injections in the corrugator muscle blocks activity in the muscle and reduces frown lines. According to simulation theory, not being able to frown should increase the difficulty of simulating sadness and anger. Remarkably, that is just what was found: Reading of sad and angry sentences was slowed after Botox injections in the corrugator muscle, but reading of happy sentences was unaffected.

One objection to simulation theory is that much of human knowledge is abstract, and hence cannot be captured by bodily systems of perception, action, and emotion. However, this objection is incorrect in at least four ways. First, there are data demonstrating that at least some abstract ideas are closely related to bodily systems. For example, Glenberg et al. (2008) demonstrated that understanding of sentences describing the transfer of abstract information (e.g., "Anna delegates the responsibilities to you") activates the motor system to the same extent as the understanding of sentences describing the transfer of concrete objects (see Santana & de Vega, 2011, for additional examples). As another example, Kousta, Vigliocco, Vinson, Andrews, & Del Campo (2010) have shown that many abstract ideas have a significant grounding in the emotional system.

Second, Barsalou (1999) has proposed that some abstract ideas can be understood as embodied relations or processes. Consider, for example, truth. The notion of truth often requires a comparison between a situation and a description of the situation. Thus, to ask, "Is it true that it is raining outside?" requires a comparison between the perception of the weather outside with a simulation of the sentence, "It is raining outside." To the extent that the perceptual reality matches the simulation, the sentence is true.

Third, Lakoff (e.g., Lakoff, 1987) has adduced a tremendous amount of linguistic evidence that people think (and talk about) abstract ideas through a process of metaphorical extension. For example, most of us do not have a refined idea of what a theory is. Instead, we think about it as a physical structure. Because of that, we say things such as "The theory has a strong foundation," or "The theory is built out of thin air," and "The theory collapsed from the weight of the evidence."

Finally, there is now some speculation and data demonstrating that even the rules of syntax are not abstract in the sense of divorced from neural systems of perception, action and emotion. Instead, Glenberg & Gallese (2011) have proposed that syntax emerges from action control.

In summary, a strong case can be made for simulation theory of language comprehension. That is, we understand language by using neural and bodily systems ordinarily used for control of perception, action, and emotion to simulate the situation described by the language. Thus, we understand language much like we understand situations: in terms of the actions the situation, or described situation, affords.

Reading comprehension

The simulation account of reading comprehension is both similar to and different from the account of oral language comprehension. It is similar in two respects. First, comprehension of written text also requires a simulation of the situation described, and that simulation is based on neural systems of action, perception, and emotion. Second, for both oral language comprehension and text comprehension, the simulation arises from three processes as described by the Indexical Hypothesis (Glenberg & Robertson, 2000; Glenberg & Gallese, 2011; Kaschak & Glenberg, 2000). The first process is indexing the linguistic symbols, that is, words and phrases, to perceived objects or previous experiences. These experiences are encoded in memory as perceptual symbols (Barsalou, 1999), that is, as aspects of the perceptual, action, and emotional neural activity engendered during previous interactions with the objects. The second process is deriving affordances from the objects, that is, how it is possible to interact with the objects given the physical nature of the objects, limits of the perceiver's body, and cultural norms. The final process is meshing or integrating the affordances according to syntax. That is, the affordances are integrated so as to create a simulation of the "who does what to whom" specified by the syntax.

Nonetheless, comprehension of written and oral language differ. One difference is in how oral and written language are learned and the consequences of that learning for later performance. In learning an oral language, the indexing of symbols (spoken words and phrases) to objects is frequent and immediate. For example, a mother will say, "here is your bottle," and give the baby its bottle; or, a father will say, "Wave bye-bye" and gesture waving; or a sibling will say, "Get the ball," and point to the ball. From these interactions, the process of moving from the linguistic symbol to the indexed object or perceptual symbol is practiced, literally from day one, and becomes fast and automatic.

Relatedly, comprehension of oral language makes use of situated cues that are normally unavailable while reading. First, conversation is often about aspects of the immediate environment to which both interlocutors are attending. Thus, the objects to which words must be indexed are primed and easily accessed in the perceptual field. Second, oral language makes use of gestures (Beilock & Goldin-Meadow, 2010; Goldin-Meadow & Beilock, 2010; Hostetter & Alibali, 2008) that help to disambiguate the language and aid in indexing. Third, oral language makes use of prosody to aide in foregrounding information and conducting the correct syntactic analysis. And finally, oral language makes exquisite use of emotional information expressed in tone, prosody, facial gesture, and full-body gestures to help convey the message. For example, in talking about a sad event one may talk softly and slowly while expressing sadness on the face and with a slumped body. All of these cues to meaning are missing when reading.

Consider learning to read. Instead of frequent and immediate indexing of linguistic symbols to objects and experiences, when learning to read indexing is slow, unreliable, and rarely practiced. The beginning reader must first deal with learning the alphabet and the alphabetic principal. Then, written words can be decoded and pronounced either through a process of phonological composition or arduous memorization of individual word-forms. Even when successfully decoded and pronounced, however, the pronunciation of the written word is unlikely to be as fluent and prosodic as in speech. Thus, the word pronounced during non-fluent reading is a poor retrieval cue for meaning, that is, the non-fluent pronunciation is difficult to index to the appropriate perceptual symbol. Furthermore, when reading, the relation between the written word and object is rarely demonstrated. For example, on successfully decoding "dog," it is rare that there is a literal dog in the environment. Even when the text might have a picture of a dog, reference to the picture is haphazard. Consequently, the link between the symbol and the embodied experiences is more tenuous in the case of reading than in the case of oral language learning. Finally, many of the cues to appropriate simulation (e.g., emotional prosody) available in oral language are missing when reading. For those children who fail to make the link between the written word and the embodied experience, reading becomes a boring exercise in word-calling that rarely results in meaning.

The Moved by Reading intervention

How can a child be taught to simulate written language? Emphasizing fluency is unlikely to provide much help. Yes, to the extent that the child's pronunciations become fluent and prosodic, as in oral communication, then those pronunciations may be effective in tapping embodied experiences. However, developing this level of fluency when there is not much meaning sounds like torture. Also, it is hard to imagine how text can be read with natural prosody before meaning is easily available.

Another strategy might be to exhort the child to think about the meaning: Who has done what to whom? Why? How does this text relate to a previous text? Use your background knowledge! Create pictures/movies in your head! But, all of these exhortations presume that the child has access to meaning and is just too lazy to use it. What if the child is struggling to derive meaning from the written text?

Moved by Reading is a two-stage reading comprehension intervention designed to overcome these problems. Children read stories that relate to a particular scenario. For example, one scenario consists of stories that take place on a farm and involve animals, objects such as a tractor, and the farmer (see Figure 1). Another scenario consists of stories that take place in a house involving a mother, father, baby, and various props such as beds and a stroller. While reading these stories, the children have available either toys corresponding to the objects and characters in the scenario, or images of the toys on a computer monitor. The first stage of *Moved by Reading* is called physical manipulation (PM). During PM, children read aloud one sentence, and then move the toys or the images to simulate the content of the sentence. For example, if the sentence is "The farmer drives the tractor to the barn," then the child is to place the toy farmer into the tractor and move the tractor to the barn.

a)

Halloween

It is almost Halloween.

Ben needs to set up his pumpkins.

Ben hooks the cart to the tractor.

Ben drives the tractor to the pumpkins.

He puts the pumpkins into the cart.

He drives the tractor to the barn.

He sets the pumpkins next to the barn.

Now, Ben is ready for Halloween.

b)

Halloween

It is almost Halloween.

Ben needs to set up his pumpkins.

Ben hooks the cart to the tractor.

Ben drives the tractor to the pumpkins.

He puts the pumpkins into the cart.

He drives the tractor to the barn.

He sets the pumpkins next to the barn.

Now, Ben is ready for Halloween.

c)

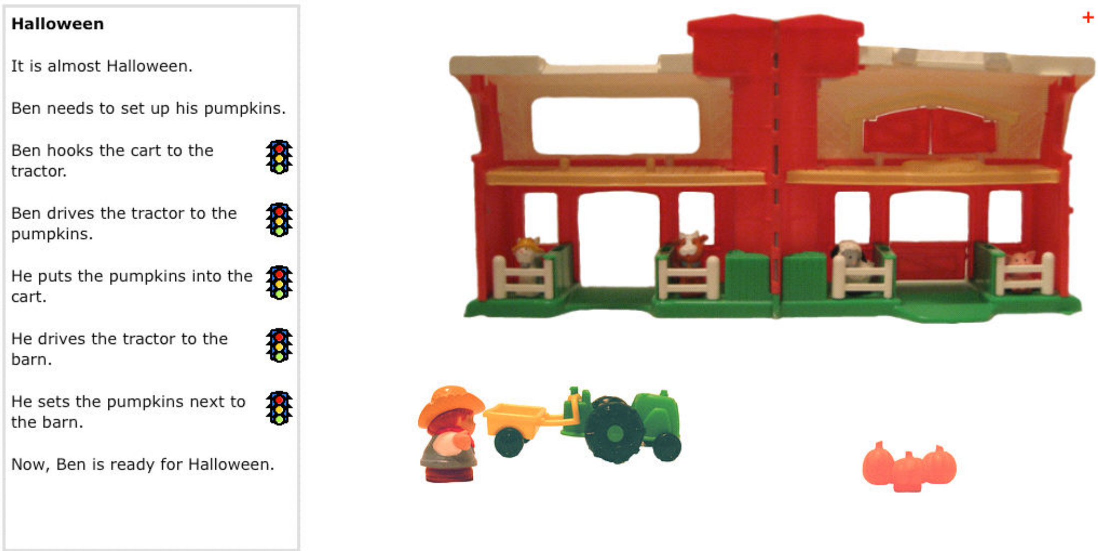


Figure 1. *(a) Screen shot before reading, "Ben hooks the cart to the tractor." (b) Screen shot midway through manipulating for "Ben hooks the cart to the tractor." (c) Screen shot after successfully manipulating for "Ben hooks the cart to the tractor." The green traffic light is the signal for the child to manipulate the toys to correspond to the sentence. Reprinted with permission from Glenberg et al. (2009)*

PM is designed to increase comprehension for the following reasons. First, the child must index the major content words to objects (or their images). This indexing can be done on a word-by-word basis, that is, it does not require understanding of the whole sentence. Because the objects are physically present, they both prime the pronunciation of the words and help to constrain the objects to which the words can be indexed. Second, the child must act out the sentence, that is, the child must physically instantiate the syntax of the sentence, the who does what to whom, in his or her own actions. The constraints of the situation (e.g., that a tractor can move easily, but a barn cannot) prime and constrain the actions that the child takes in producing the simulation. Finally, PM demonstrates for the child how written texts can be meaningful and how to uncover that meaning.

PM results in large increases in comprehension (as reviewed below), but if children always needed to physically manipulate to understand written text, it would not be very practical. Fortunately, after using PM, children can be relatively easily transferred to Imagined Manipulation (IM). The children are told to imagine moving the objects or images just as they did with PM, but now the objects and images are not physically present. Instead, the child indexes the written words to the perceptual symbols of those objects.

Visualizing or imaging content has a well-researched history in the domain of text comprehension (Bell 1986, Paivio 1986, Sadoski & Paivio, 2001). There are several important differences between earlier research and *Moved by Reading,* however. First, the instruction during IM ("imagine moving the characters like you just did") is clearer than the instruction given in many experiments investigating imagery ("create pictures in your head"). This clarity arises not from the words, but from the fact that the child has just manipulated using PM. Thus, the child can understand the IM instruction *itself* by indexing that instruction to the embodied experiences created during PM. Second, IM, in contrast to imagery instructions, is likely to engender a significant motor component in addition to visual imagery (see also, Varley, Levin, Severson, & Wolff, 1974 and Wolff & Levin, 1972). Including a motor component increases the range of information encoded. In addition, given the role of motor cortex in prediction (e.g., Bubic, Von Cramon, & Schubotz, 2010), eliciting motor activity should enhance predictive processing while reading. Thus, reading using IM enhances comprehension by encouraging indexing, by encoding multiple sources of information, and by enhancing predictive processing. When children become accomplished at IM, they become accomplished readers.

Does Moved by Reading enhance reading comprehension?

Yes, and often dramatically. Glenberg, Gutierrez, Levin, Japuntich, and Kaschak (2004) implemented an early version of *Moved by Reading* with children in the first and second grades. Working one-on-one with the experimenter, these children literally manipulated toys to simulate sentence content during PM. Performance was compared to children in a control group who read and re-read the same texts. Children in the control condition also had the toys visible, but these children did not manipulate the toys. Consider first recall of the action sentences. Children who used PM recalled 62% and children in the control condition recalled 29%, and the effect size (Cohen's d, the number of standard deviations between the two means) was 1.39[1]. The advantage was also found when the children were tested on their ability to correctly answer inferences based on the text, $d = .81$. Children who had used PM and were transferred to IM outperformed children in the control condition both on correct recall ($d = 1.87$) and correct inference answers ($d = 1.50$). On the third day of the experiment, the children were not given any special instructions for reading. Nonetheless, the children who had practiced PM and IM on the previous days outperformed the children in the control condition on recall ($d = 1.23$) and in answering inferences ($d = .95$, although due to a small sample size, the inference effect was not significant at the .05 level).

Recently, we implemented *Moved by Reading* as a web-based system (Glenberg, Goldberg, & Zhu, 2009; Glenberg, Willford, Gibson, Goldberg, & Zhu, 2011). When using the computer, children manipulated images on the computer screen rather than directly manipulating toys. Manipulation of the images produced benefits ($d = 1.16$) relative to re-reading comparable to those found when manipulating the toys.

Must Moved by Reading be implemented one-on-one?

Not at all. Glenberg, Brown, and Levin (2007a) implemented *Moved by Reading* with three-child reading groups. During PM, one child would read and manipulate, and then the next child would read and manipulate the next sentence, and so on. Over all, the groups that engaged in PM were much more accurate in answering comprehension questions than were children who read and reread the texts, $d = 1.72$. Interestingly, the effect was found both for questions that tapped understanding of sentences a particular child manipulated ($d = 1.26$) and for questions that tapped understanding of sentences manipulated by other children in the group ($d = 1.86$). At first glance, this result was surprising, although on reflection there are several possible explanations. First, watching another manipulate generates a vision-based memory in addition to any language-based memory. Second, watching another act will stimulate the observer's mirror neuron system (Glenberg, 2011; Rizzolatti & Craighero, 2004) and lead to neural activity in the observer's motor system that is substantially similar to the activity when the person is engaged in literal action.

[1] Values of d around .2 are considered a small effect, d around .5 is a medium effect, and $d > .8$ is considered a large effect.

Recently, we were able to implement *Moved by Reading* for a whole classroom of students using the web-based versions of PM and IM (Glenberg et al., 2011). In this experiment, the children read to solve mathematical story problems (which will be described in more detail shortly). When measuring problem solving, children who used PM were more accurate than children who read (d = 1.19). Similarly, when those children went on to use IM while reading the story problems, they were more successful than the students who simply read the texts (d = .90).

Are there long-term benefits of Moved by Reading?

Here the data are not as secure as one would like. In Glenberg et al. (2004), there was about a week's delay between training in PM and training in IM, and another week intervened before testing with no further instruction. This final test did show a positive effect of *Moved by Reading* training, d = 1.11. For older children (third and fourth grades), Glenberg, Jaworski, Rischal, and Levin (2007b) found that minimal experience with *Moved by Reading* facilitated performance some three weeks later (d = .48). It is likely that more extensive training with *Moved by Reading* will produce larger and longer-lasting effects.

Is Moved by Reading effective for special populations?

Yes. Perhaps the most dramatic demonstration is provided by Marley, Levin, and Glenberg (2007) who investigated the listening comprehension of learning-disabled Native American children. The children were randomly assigned to three conditions. In the PM condition, the children manipulated after listening to the experimenter read a sentence; in the visual condition, the children heard the experimenter read the sentence and then watched him manipulate; in the free study condition, the children listened to the experimenter and were instructed to think about each sentence in the pause following the sentence (equivalent to the time needed to manipulate). Children in the PM and visual conditions outperformed children in the free study condition with $ds > 1$ for free recall of propositions, objects, and actions (but not locations), as well as cued recall. There were minor differences in favor of the PM condition relative to the visual condition (d = .32), but for the most part, these two conditions were similar. The similar performance in the PM and visual conditions probably reflects the operation of the mirror neuron system as discussed above in the context of the data from the three-person reading groups (Glenberg et al., 2007).

One way in which the data on listening comprehension differed from the usual pattern with *Moved by Reading* is that the effects observed with PM did not transfer to an IM condition. In retrospect, there are two reasons why IM was not effective in this experiment. First, the children were not given any scaffolded instruction about how to use IM [the experiments reported in (Glenberg et al., 2004) demonstrated the importance of this type of scaffolding]. Second, the children were instructed to "close your eyes and make pictures in your head." This instruction does not connect as well to PM as the standard IM instruction to "imagine manipulating the toys just as you did before."

Non-disabled Native American children were participants in research reported by Marley, Levin, and Glenberg (2010). With third-grade children, the expected *Moved by Reading* results were found. Namely, PM resulted in better free recall of stories than a reread condition, d = 1.45, and the same was true for IM (although using the instruction to "make pictures in your head"), d = 1.09. For children in the second grade, PM was effective, d = .84, but IM was not (d = .40, but the difference was not significant at the .05 criterion). These data indicate that younger children probably need more scaffolding to implement IM than do older children (see also, Marley, Szabo, Levin, & Glenberg, 2011)

Does Moved by Reading training transfer?

There are two types of transfer to consider. The first is when a child receives *Moved by Reading* training with stories from one scenario (e.g., the Farm scenario) and is asked to apply IM to stories from another scenario. In fact, the data reported above from Glenberg et al. (2007b) were collected when children were trained and tested on different scenarios. Similarly, in Glenberg et al. (2004), during the final session in which children read with no further instruction the texts were from a new scenario.

Perhaps a more interesting case of transfer was reported by Glenberg et al. (2011). The major question addressed was whether *Moved by Reading* could be considered a fundamental reading strategy, that is, one that could be used when reading in any domain. To begin to answer this question, third- and fourth-grade children were taught and practiced *Moved by Reading* while reading in narrative and expository-like domains. Then, the children were asked to solve mathematical story problems written using characters and situations from the narrative and expository-like domains. The children were not given any special instruction for how to use *Moved by Reading* in a story problem context (in contrast to Glenberg, et al., 2007b). Will helping children to understand the stories help them to solve the math problems?

Table 1. *A mathematical story problem from Glenberg et al. (2011)*

Ray enters a talent contest at school. He uses his telekinesis to perform a magic trick. It is his favorite: levitation. When objects levitate they float in the air.
Ray is among 7 contestants sitting in front of the stage waiting to perform. He is number 4. His number is called and he leaps to the stage.
There are 2 objects on the stage: a computer and a dumbbell. He surprises the audience by making the computer levitate. It weighs 22 pounds.
He hears a gasp from the audience. He then levitates the dumbbell. It weighs 55 pounds.
How many pounds in total has he lifted?
Please show all of your work.

Note: Each of the first three paragraphs ended in a green dot used as a cue to perform PM or IM (Moved by Reading condition) or an indication of important information for the Control condition.

We used two primary measures of problem-solving performance. The first was simply whether the numerical answer was correct. The second reflects a finding that Verschaffel, Van Dooren, Greer, and Mukhopadhyah (2010) describe as suspension of sense-making. That is, when some children approach story problems, they appear to give up on making sense of the story and simply try to combine numbers to do the math. To detect the suspension of sense-making, each of the stories had story-relevant numerical information that was irrelevant to solving the problem. An example is given in Table 1. In the text, the number 7 (the number of contestants) and the number 4 (Ray's number) are story-relevant, but are irrelevant to solving the story problem (how many pounds Ray lifted). If *Moved by Reading* helps children to understand the story in a way that constrains problem solving, then there should be a reduction in misuse of the irrelevant information.

Focusing on the data from the third day of the experiment in which the children used IM during story problem solving, children who had been trained in *Moved by Reading* correctly solved 44% of the problems, whereas the children in the control condition solved 33% (d = .59). In the *Moved by Reading* condition, children misused irrelevant information 38% of the time, whereas in the control condition, the children misused the irrelevant information 61% of the time (d = .78).

These data strongly support the possibility that *Moved by Reading* teaches a fundamental reading comprehension strategy. That is, even though *Moved by Reading* was taught in the context of narrative comprehension, the same strategies supported comprehension that can be used to solve mathematical story problems. Clearly, there are many other types of text genres and reading goals, and there is no guarantee that *Moved by Reading* will be useful in all of them. On the other hand, if the underlying embodied account of comprehension is correct, then it seems likely that *Moved by Reading* is a fundamental strategy that will apply across many domains and tasks.

Moved by Reading and reading in abstract domains

At this point, the skeptical reader may be thinking: "*Moved by Reading* has been shown to be effective with children reading texts appropriate for children. However, adult reading concerns abstract topics such as government or physics in which there is nothing to manipulate using PM or IM. What would it mean to apply *Moved by Reading* in an abstract context?"

In fact, I think that it is likely that many (if not all) abstract topics are understood using the sort of embodied processes encouraged by *Moved by Reading*. As one example, consider reading a text in physics about centripetal force. When an object is in circular motion, then the force on the object acting toward the center of the circle is given by $F = mv^2/r$, where m is the mass of the object, v is the velocity, r is the radius of the circle, and F is centripetal force. How could information such as this be embodied?

A skilled writer will do something like the following to help the reader understand the equation. "Imagine that you are on roller skates in a parking lot. To stop, you grab a post, and as you fly by, you start to spin around the post. That spinning is circular motion, and the force that you feel in your arms is centripetal force, that is, the force causing the circular motion. The speed of your skating before grabbing the post (v) will affect the centripetal force you that feel in your arms. If you are skating fast, then you will be jerked more vigorously when you grab the post than when you are skating slowly. That is the v^2 part of the equation: The faster you go, the greater the centripetal force once you grab the post (and the more it will hurt). Now imagine that you are wearing a heavy backpack (thus you have greater mass), but that you are skating just as fast as before. Will the force that you feel in your arms when you grab the post be greater or less than without the backpack? In fact, the m part of the equation indicates that the force will be greater: If you are more massive, then it is going to hurt more to grab the post than if you were not wearing any backpack. Finally, imagine that instead of grabbing the post with your hands that you have a rope with a loop, and you lasso the post with the loop while you hold onto the other end of the rope. If the rope is short, then you will be whipped around the post in a tight circle, whereas if the rope is long, your path around the post will be a more leisurely, large circle. In which case will you feel more strain (centripetal force) on the rope and your arms? According to the equation, the radius of the circle (r) acts as a divisor so that the longer the rope, the less the force. You can get a feel for this by thinking about how much centripetal force you will feel while whipping around the post on a short rope compared to the more leisurely drift on a long rope."

Thus the skilled writer helps the reader to index the abstract symbols (F, m, v, and r) to embodied experiences. The imagery that a reader experiences in reading an example such of this is exactly the reader's use of IM.

Of course, not all writers are skilled, and in that case it is up to the reader to find a way to index the symbols to appropriate experiences. When reading in a new domain, such as physics, that can be a daunting task. And that is why so few readers understand centripetal force and other scientific concepts.

Conclusions

How is reading comprehension embodied? When understanding text, the words and phrases are indexed to embodied (that is, perception, action, and emotion) experiences to create a simulation of the content of the text. Because this simulation is in a format intimately related to the body, it can be used to guide action. And why does that matter? By understanding the embodied nature of language comprehension, we can create successful educational interventions such as *Moved by Reading.*

• • •

Acknowledgment

This article was written while the author was partially supported by NSF grant DRK-12 1020367. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the funding agencies.

Arthur Glenberg is a Professor of Psychology at Arizona State University, an ASU Learning Sciences Institute Senior Scientist, and Professor Emeritus at the University of Wisconsin-Madison. He earned his bachelor's degree in Psychology from Miami University (the real one in Oxford, Ohio) in 1970 and his Ph.D. from the University of Michigan in 1974. Glenberg's research has focused on memory and language comprehension. Beginning in the middle 1990s, he began to develop the embodied approach to cognition. Within that framework, his research has focused on language, reading comprehension interventions (Moved by Reading), and mirror neurons. His Harzing H-index is 40 with 18 papers with 100 or more citations. Glenberg was an associate editor of the Journal of Experimental Psychology: Human, Learning, and Memory, the author of an elementary statistics textbook, Learning From Data, and the co-editor of Symbols, Embodiment and Meaning. Currently, he serves on the editorial boards of five journals, and he is an elected fellow of the Association of Psychological Science, Society for Text & Discourse, and Society of Experimental Psychologists.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences*, *22*(04), 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*, 617–645.

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture Changes Thought by Grounding It in Action. *Psychological Science*, *21*(11), 1605-1610. doi:10.1177/0956797610385353

Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, *4*.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, *61*(6), 825-850. doi:10.1080/17470210701623605

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *The Multiple Functions of Sensory-Motor Representations*, *22*(3/4), 455.

Glenberg, A. M. (2007). Language and action: creating sensible combinations of ideas. In G. Gaskell (Ed.) The Oxford handbook of psycholinguistics (pp.361-370). Oxford, UK: Oxford University Press.

Glenberg, A. M. (2011). Introduction to the Mirror Neuron Forum. *Perspectives on Psychological Science*, *6*(4), 363-368. doi:10.1177/1745691611412386

Glenberg, A. M, Brown, M., & Levin, J. R. (2007). Enhancing comprehension in small reading groups using a manipulation strategy. *Contemporary Educational Psychology*, *32*(3), 389–399.

Glenberg, A. M., & Gallese, V. (2011). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex*. doi:10.1016/ j.cortex. 2011. 04.010

Glenberg, A. M., Goldberg, A. B., & Zhu, X. (2009). Improving early reading comprehension using embodied CAI. *Instructional Science*, *39*(1), 27-39. doi:10.1007/s11251-009-9096-7

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, *96*(3), 424–436.

Glenberg, A. M., Jaworski, B., Rischal, M., & Levin, J. R. (2007). What brains are for: Action, meaning, and reading comprehension. *Reading comprehension strategies: Theories, interventions, and technologies*, 221–240.

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*(3), 379–401.

Glenberg, A., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., & Buccino, G. (2008). Processing abstract language modulates motor system activity. *The Quarterly Journal of Experimental Psychology*, *61*(6), 905-919. doi:10.1080/ 17470210701625550

Glenberg, A. M., Willford, J., Gibson, B., Goldberg, A., & Zhu, X. (2011). Improving Reading to Improve Math. *Scientific Studies of Reading*, 1-25. doi:10.1080/ 10888438. 2011.564245

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's Influence on Thought: The Case of Gesture. *Perspectives on Psychological Science*, *5*(6), 664-674. doi:10.1177/ 1745691610388764

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307.

Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., & Davidson, R. J. (2010). Cosmetic Use of Botulinum Toxin-A Affects Processing of Emotional Language. *Psychological Science*, *21*(7), 895-900. doi:10.1177/0956797610374742

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495-514. doi:10.3758/PBR.15.3.495

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*(3), 508–529.

Kiefer, M., & Pulvermüller, F. (2011). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*. doi:10.1016/j.cortex.2011.04.006

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2010). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*. doi:10.1037/a0021446

Marley, S. C, Levin, J. R, & Glenberg, A. M. (2007). Improving Native American children's listening comprehension through concrete representations. *Contemporary Educational Psychology*, *32*(3), 537–550.

Marley, S. C., Szabo, Z., Levin, J. R., & Glenberg, A. M. (2011). Investigation of an Activity-Based Text-Processing Strategy in Mixed-Age Child Dyads. *The Journal of Experimental Education*, *79*(3), 340-360. doi:10.1080/00220973.2010.483697

Marley, S., Levin, J., & Glenberg, A. (2010). What Cognitive Benefits Does an Activity-Based Reading Strategy Afford Young Native American Readers? *The Journal of Experimental Education*, *78*(3), 395-417. doi:10.1080/00220970903548061

Rizzolatti, G., & Craighero, L. (2004). The mirror neuron system. *Annual Review of Neuroscience*, *27*(1), 169-192. doi:10.1146/annurev.neuro.27.070203.144230

Rueschemeyer, S.-A., Glenberg, A. M., Kaschak, M. P., Mueller, K., & Friederici, A. D. (2010). Top-Down and Bottom-Up Contributions to Understanding Sentences Describing Objects in Motion. *Frontiers in Psychology*, *1*. doi:10.3389/ fpsyg.2010. 00183

Santana, E., & de Vega, M. (2011). Metaphors are Embodied, and so are Their Literal Counterparts. *Frontiers in Psychology*, *2*. doi:10.3389/fpsyg.2011.00090

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1-11.

International Electronic Journal of Elementary Education, 2011, 4(1), 19-31.

Improving middle and high school students' comprehension of science texts

Brandi E. JOHNSON
Georgia State University, United States

Karen M. ZABRUCKY *
Georgia State University, United States

Abstract

Throughout the United States, many middle and high school students struggle to comprehend science texts for a variety of reasons. Science texts are frequently boring, focused on isolated facts, present too many new concepts at once, and lack the clarity and organization known to improve comprehension. Compounding the problem is that many adolescent readers do not possess effective comprehension strategies, particularly for difficult expository science texts. Some researchers have suggested changing the characteristics of science texts to better assist adolescent readers with understanding, while others have focused on changing the strategies of adolescent readers. In the current paper, we review the literature on selected strategy instruction programs used to improve science text comprehension in middle and high school students and suggest avenues for future research.

Keywords: reading comprehension, comprehension of science texts

Introduction

Reading comprehension involves a set of multifaceted and interconnected skills allowing students to accurately process and understand text information during reading (Zimmerman, Gerson, Monroe, & Kearney, 2007). The processes involved in reading comprehension include, in part, focusing on relevant and important information from a passage and making connections between that information and prior knowledge. But students must also understand the meaning of words as well as integrate the many internal connections among important and relevant pieces of information within a passage (Baker, 1985; Cook & Mayer, 1988). Several researchers (e.g., Cook & Mayer, Magliano, Todaro, Millis, & Wiemer-Hastings, 2005) have expanded upon the typical definition of comprehension by

* Karen M. Zabrucky, Department of Educational Psychology and Special Education, P.O. Box 3979, Georgia State University, Atlanta, GA , 30302-3979, United States. E-mail: zabrucky@gsu.edu

ISSN:1307-9298

suggesting that deeper comprehension results from students purposefully trying to reach a coherent understanding of what a text is about. When reading difficult texts, skilled readers use a variety of comprehension strategies to build deeper meaning.

Within the United States, large proportions of middle and high school students struggle to read and understand content area textbooks. It is not uncommon, within some schools, for 75-80% of the students in a significant number of classes to be unable to successfully read textbooks (Carnine & Carnine, 2004). The epidemic has become so great that the state of California designated a new category for such students, labeled "struggling readers." According to Bhattacharya (2006), students must accurately and fluently read passages containing extensive vocabulary with multiple syllables to successfully comprehend content-area texts. Students tend to struggle in particular with comprehension of science texts. Even if they can decode, read, and understand the words in the texts, students have problems making the words make sense. The words appear as a string of known and unknown words rather than a message that is coherent, comprehensible, and learnable for students (Best, Rowe, Ozuru, & McNamara, 2005).

Factors Contributing to Students' Difficulties Comprehending Science Texts
Science Texts: Content and Structure Issues

Several factors may contribute to students' poor lack of understanding of science texts. The texts themselves may cause problems because science texts are frequently inaccurate, focused on isolated facts, boring, and poorly organized (Chambliss & Calfee, 1989). Carnine and Carnine (2004) further criticized science texts by stating that such texts contain too many vocabulary concepts, present too many ideas at once, lack clarity, and fail to transmit science knowledge. Several of these characteristics would appear to fly in the face of fundamental processes that affect ease of text comprehension (Kintsch & van Dijk, 1978).

It is very likely that students' comprehension skills contribute greatly to their struggles with science texts, which may be too demanding for students' skill levels. According to Cook and Mayer (1988), students may be unaware of the underlying structure of passages within a science text. The construction integration (CI) model of text comprehension emphasizes that domain-knowledge drives text comprehension and, thus, students with limited existing knowledge of science concepts will experience difficulty comprehending science texts (Best et al., 2005).

There are several approaches taken by researchers to improve students' comprehension of science tests. One approach involves changing the design of science textbooks. Chambliss and Calfee (1989) found multiple differences in science textbooks for nine-year-old students in Japan, Singapore, and the United States. In comparison with texts from the other countries, science texts from the United States were not only larger but more cluttered with information and details, resembling incoherent compilations rather than "teaching books" (p. 313). Moreover, science texts use an expository rather than narrative structure more familiar to students (Cook & Mayer, 1988).

The expository texts used in school classrooms are often low in cohesiveness and too demanding for students with little background knowledge in a particular content area. Experts who write such expository texts often inaccurately assume that students possess prior knowledge of subject matter similar to the writer's prior knowledge. As noted by Best, Floyd, and McNamara (2008) "In contrast to narrative texts, expository texts tend to place increased processing demands on the reader due to their greater structural complexity, greater informational density, and greater knowledge demands " (p. 140).

Studies show that texts with high cohesion benefit readers with less domain knowledge (Best et al., 2005). Chambliss and Calfee (1989) recommend that in large science texts, content should be organized coherently and explicitly. Authors should intertwine subject matter with student knowledge and use functional devices like introductions, transitions, and conclusions to pull text information together. Carnine and Carnine (2004) also argued that extraneous information in middle school textbooks was greatly reduced when the content was simplified and instruction focused on a few key concepts. To improve retention of text information the authors encouraged review of core concepts through the use of embedded questions throughout a text and use of discussion questions to direct class discussions related to a text.

Students' use and knowledge of relevant strategies for comprehending science texts

While some researchers interested in improving students' comprehension of science texts have focused on the issue of making textbooks more coherent, others have conducted systematic examinations of students' strategy use. In the present article, we will review research on selected programs used to improve middle and high school students' comprehension of science texts.

According to the 2000 National Assessment of Educational Progress (NAEP) in science, only 32% of the nation's 8th graders performed at or above the level of *Proficient.* Further, the number of 12th graders performing at or above the *Basic level* declined between 1996 and 2000 (Carnine & Carnine, 2004). Eighth grade students who perform at the *Proficient* level demonstrate much of the knowledge and many of the reasoning abilities essential for understanding of the Earth, physical, and life sciences at a level appropriate to grade 8, while seniors performing at the *Basic* level demonstrate some knowledge and certain reasoning abilities required for understanding of the Earth, physical, and life sciences at a level appropriate to grade 12 (National Center for Education Statistics: A Nation's Report Card Science, 2010, para. 16). Meanwhile, the 2003 NAEP in reading revealed that 26% of eighth graders could not read at the basic level, indicating that many adolescents do not understand what they read (McNamara, O'Reilly, Best, & Ozuru, 2006). Eighth-grade students performing at the *Basic* level should demonstrate a literal understanding of what they read and be able to make some interpretations. When reading texts appropriate to eighth grade, they should be able to identify specific aspects of the text that reflect overall meaning, extend the ideas in the text by making simple inferences, recognize and relate interpretations and connections among ideas in the text to personal experience, and draw conclusions based on the text (National Center for Educational Statistics: A Nation's Report Card Reading, 2010, para. 16).

The statistics regarding student performance on the science and reading NAEP show why there is a growing concern in the United States with students' ability to read, comprehend, and learn from texts, especially in the area of science. Too many middle and high school students struggle with reading and comprehending science texts. Increasing the percentage of students who can successfully comprehend science textbooks requires an improvement in students' comprehension strategies.

iSTART. One program examined by researchers to help middle and high school students learn strategies and improve comprehension of science texts involves an animated conversational agent called Interactive Strategy Trainer for Active Reading and Thinking (iSTART). Graesser, Jeon, & Dufty (2008) suggest that animated conversational agents, which actually interact with students, help students learn by holding a conversation with the students and/or modeling good pedagogy for them. Students communicate with the agents by talking to them, using a keyboard, making gestures, or using a touch panel screen or input channels. The agents communicate back with students through speech, facial expressions, gestures, posture, etc (Graesser et al., 2008).

iSTART is a web-based reading strategy program that helps students learn metacomprehension strategies that support them in developing a deeper comprehension as they read difficult science texts (Graesser, McNamara, & VanLehn, 2005). iSTART stemmed from a successful classroom intervention called Self-Explanation Reading Training (SERT) that combined self-explanation, or explaining what a sentence or portion of text means, with reading strategy training. Training resulted from empirical findings that revealed that students who can self-explain are more successful at solving problems, more likely to generate inferences, construct more coherent mental models, and develop a deeper understanding of the concepts discussed in a text (McNamara et al., 2006). Graesser et al. (2008) noted that iSTART requires students to create self-explanations of text by using the five reading strategies of monitoring comprehension, paraphrasing explicit text, making bridging inferences between the current sentence and prior text, making predictions about the subsequent text, and elaborating the text with links to what the reader already knows.

iSTART consists of three modules which include an introduction, demonstration, and practice. During the introduction, students receive information about five reading strategies from two animated students and a teacher animated conversational agent. After learning about a strategy, the students complete a multiple-choice quiz to assess their understanding of the strategy (Graesser et al., 2005). The second module, the demonstration module, identifies ways that the reading strategies can be used to self-explain expository texts. Specifically, two animated characters, Merlin (the teacher) and Genie (the student), demonstrate the use of self-explanation. The students receiving training identify and select on a computer screen the strategy Genie used to self-explain a science text. Merlin then provides verbal feedback to Genie about the quality of his self-explanation. Finally in the practice module, the students receiving training type their own self-explanations for science texts and Merlin assesses the quality of their self-explanations and provides feedback to the students (McNamara et al., 2006). Merlin may ask the students to modify self-explanations until the self-explanations reach a satisfactory level. The students must then identify the reading strategies they used in their self-explanations (Graesser et al., 2005).

McNamera et al. (2006) conducted a study to examine the effectiveness of iSTART in helping adolescent readers learn reading strategies, and improve their comprehension of science texts. Participants in the study included 39 children enrolled in a summer learning program in the Eastern United States with approximately half of the students entering the eighth grade and half of the students entering the ninth grade. All participants were administered the Metacognitive Strategy Index (MSI) at their school as a group, one week prior to training. The MSI is a 25-item multiple-choice questionnaire that measures knowledge of metacognitive reading strategies (McNamera et al., 2006). During the training session, the control group, consisting of approximately one half of the students, was provided with only the initial portion of the iSTART introduction, which describes the concept of self-explanation and provides an example of a self-explanation. The experimental group received a one-hour training session for two consecutive days on all three iSTART modules. One day after training, both groups read and explained a text about heart disease. Students were required to self-explain each sentence of the text as they read it, without receiving feedback. Students then answered comprehension questions on paper about the heart disease text. Results revealed that both iSTART training and prior knowledge of reading strategies improved the quality of self-explanation and, therefore, comprehension. Students with more prior knowledge of reading strategies benefited most on bridging inference questions after iSTART training. They were able to make more bridging inferences and elaborations than students in the control condition, which allowed them to perform better on bridging inference questions. Students with limited prior knowledge of reading strategies before iSTART training learned how to develop a coherent understanding of the information presented in the text and therefore performed better than controls on text-based questions after receiving training (McNamera et al., 2006).

The results of the study by McNamera et al. (2006) were consistent with the results of a similar study in which investigators examined the effect of iSTART on students' comprehension of science texts (O'Reilly, Sinclair, & McNamara, 2004). The researchers administered three aptitude tests, the prior science knowledge test, the Gates-MacGinitie Reading Skill Test, and the MSI to thirty-eight middle school students participating in a Learning Bridge summer program. The investigators wanted to examine students' knowledge of different science domains, metacognitive reading strategies, and level of standardized reading comprehension. One week later, half of the students received two consecutive days of one-hour sessions of training over the iSTART introduction, demonstration, and practice modules. Students self-explained one text about thunderstorms and one text about coal during the practice module by typing their explanations into a computer. Remaining students served as a control group and received a description and examples of self-explanation, but did not receive iSTART training or practice with the system. Similar to the study by McNamera et al., (2006) students in the iSTART and control groups then read and self-explained each sentence of a text on heart disease. Students did not receive feedback from iSTART, but did answer comprehension questions about the heart disease text on paper (O'Reilly et al., 2004). Results indicated that iSTART training improved comprehension of science texts, but had different effects on students with high knowledge of reading strategies versus students with low knowledge of reading strategies. Students with high knowledge of reading strategies performed better on bridging questions after iSTART training as compared to the control group. Students with low knowledge of reading strategies performed better on text-based questions than the control group (O'Reilly et al., 2004).

Magliano et al. (2005) supported findings of O'Reilly et al. (2004) and McNamera et al. (2006) by examining changes in reading strategies that occur in readers of different skill levels as a function of iSTART training. The Magliano et al. study took place across four sessions within approximately one month. Each session lasted about an hour and a half. Magliano et al. administered the Nelson-Denny test, a domain specific test, and a general science knowledge test to fifty-three college students enrolled in an introductory psychology course. One week later, the students participated in a pre-iSTART session using Microsoft Excel, in which they were told to type self-explanations of each sentence embedded in two- to five-sentence scientific texts as they appeared on the computer screen. Students were told to self-explain by producing whatever thoughts immediately came to mind regarding their understanding of a sentence in the context of a text. Students then took a short-answer comprehension test after reading both texts.

During the next week, students engaged in the iSTART computerized system at their own pace. The post-iSTART session occurred one to two weeks after the third session. The post-iSTART session was similar to the pre-iSTART session except that during the post-iSTART session, students were explicitly instructed to practice iSTART reading strategies when producing self-explanations. The researchers provided the participants with a list of the strategies and the definitions of the strategies as a reminder (Magliano et al., 2005). The results of the study indicated that only skilled readers engaged in more global processing after iSTART training, but both skilled and less-skilled readers increased their strategy use and produced more relevant self-explanations after iSTART than before. Consistent with prior research, after iSTART training less skilled readers improved their performance on text-based questions, but not bridging questions. However, skilled readers improved their performance on bridging questions.

PALS. Peer Assisted Learning Strategies (PALS) is an alternative program explored by researchers to improve adolescents' reading comprehension of science texts. PALS is a reading comprehension strategy program based on a class wide peer tutoring (CWPT) model. CWPT is a system in which all class members are organized in tutor-tutee pairs and work together rather than independently or in small groups (Calhoun, 2005). It provides students with an increase in practice opportunities, immediate error correction, pacing, content coverage, high mastery levels, and immediate feedback Researchers have found that students participating in CWPT outperform students in control classrooms in reading, spelling and mathematics, at both the elementary and secondary levels (McMaster, Fuchs, & Fuchs, 2006) and that students remain engaged in PALS nearly 100% of the time (Calhoun, 2005).

Given that inquiry plays a significant role in the scientific process and requires students to be engaged, teachers must find ways to encourage student engagement in science classes (Kroeger, Burton, & Preston, 2009). Peer-mediated instructional practices like CWPT and PALS may support science learning by keeping students engaged and PALS has been shown to improve students' comprehension of science texts while keeping them engaged (Kroeger et al., 2009). Approximately fifteen years of pilot studies, component analyses, and large-scale experiments conducted within classrooms have indicated that PALS improves the reading achievement of low, medium, and high achieving students (McMaster et al., 2006). In fact, PALS earned *Best Practice* status from the U.S. Department of Education Program Effectiveness Panel (McMaster et al., 2006).

According to Calhoun (2005) students participate in three essential reading comprehension activities while reading aloud during PALS. The activities include Partner Reading with Retell, Paragraph Shrinking, and Prediction Relay. All three activities are designed to provide students with practice in reviewing, sequencing, stating main ideas, summarizing main ideas, and predicting outcomes. For each activity, the higher-performing student reads the text first, followed by a lower-performing student who acts as the first Coach. The readings come from texts that are at an appropriate level for the lower-performing student in each pair (McMaster et al., 2006).

During Partner Reading with Retell each student reads aloud from connected text for five minutes. The higher-performing reader reads a passage in a text. If the reader makes an error while reading, the Coach asks the reader to stop reading and sees if the reader can figure out the word. If so, the reader says the word and continues reading. If not, the Coach tells the reader the word. The reader then repeats the word and rereads the sentence. The students then switch roles. After both students read, the lower-performing student retells the sequence of events read for two minutes. Students earn a point for each sentence read correctly and ten points for the retell (McMaster et al., 2006).

Paragraph Shrinking constitutes the second PALS activity. During Paragraph Shrinking, the students continue to read aloud, but they stop at the end of each paragraph to identify the main idea. The Coach asks the reader to identify who or what the paragraph is talking about and the most important thing about the "who" or "what." The reader then condenses the information into ten words or less (McMaster et al., 2006). The reader earns one point for identifying who or what the paragraph is talking about, one point for stating the most important thing, and one point for stating the main idea in ten words or less (McMaster et al., 2006). After one reader shrinks a paragraph, then the roles of Coach and reader reverse for students.

The final PALS activity is Prediction Relay. Prediction Relay consists of the reader making a prediction about what will happen on the next half page to be read, reading the half page aloud, confirming or disconfirming the prediction, and summarizing the main idea (McMaster et al., 2006). The tutor can disagree with a prediction and ask the reader to make another one. During Prediction Relay, students earn points for making reasonable predictions, reading each half page, accurately confirming or disconfirming the prediction, and identifying the main idea in ten words or less (McMaster et al., 2006).

Fuchs, Fuchs, and Kazdan (1999) provided evidence that peer-assisted learning can have a substantially positive effect on struggling high school students' reading comprehension. Their study consisted of 102 students divided into nine control and nine comparison remedial or special education classrooms within 10 high schools in a southeastern school district. Teachers in the control classrooms provided reading instruction two to three times a week over 16 weeks, using conventional reading programs without peer-assisted learning. Teachers in comparison classrooms provided the three PALS activities in their classrooms for the same amount of time.

Although the researchers randomly assigned teachers to control or comparison groups, the teachers identified the students whose data would be included in the analysis. These students all read between grade levels two and six. While the results of pre and post assessments on the Comprehension Reading Assessment Battery were not statistically significant between control and comparison groups, the researchers classified the effect of peer-assisted learning in this study as important because the effect size was greater than 0.25. Furthermore, results indicated that the average student in the control group would have gained at least 13 percentile points in achievement if the student had received PALS training.

Typically, PALS is used to supplement existing reading programs (Calhoun, 2005). Calhoun examined the combined effects of Linguistics Skills Training (LST) and PALS on the reading skill acquisition of middle school students with reading disabilities (RD). LST encompasses an age-appropriate, peer-mediated phonological skill program that uses an explicit linguistic signaling and coding system that enables students to identify the sounds of letters or letter clusters. In this study, four teachers from two middle schools in the southwest participated. Each instructor taught language arts to students with RD in a self-contained classroom. Calhoun's sample consisted of thirty-two sixth graders, five seventh graders, and one eighth grader. All students read at least three grade levels below their expected reading level. Teachers incorporated PALS twice a week with their students and LST three days a week. After 31 weeks of intervention using LST/PALS, students significantly increased their reading comprehension skills compared to a control group that used the widely used remedial reading program Saxon Phonics Intervention. LST/PALS was also effective in teaching students phonological skills even though reading fluency did not improve (Calhoun, 2005).

While most investigators have examined the use of PALS in reading narrative texts, Kroeger et al., 2009, looked at the use of PALS with expository science texts. Researchers used an adapted version of PALS, PALScience, in two classrooms with approximately 28 students in each classroom. The study took place in a large Midwestern suburban middle school. PALScience incorporates the reciprocal peer tutoring and Paragraph Shrinking activities consistent with PALS literature, but not the Prediction Relay activity.

The two teacher-researchers involved in the study chose a single-subject withdrawal design to measure intervention effects, allowing them to focus on the individual changes of students. Initially, all students were trained on the use of helping skills, catching mistakes, making positive comments, and the use of self-monitoring skills to be effective tutor-tutee partners (Kroeger et al., 2009). In the baseline and withdrawal phases of the study, students read science passages with no peer assistance. Passages were selections from seventh-grade classroom science texts that ranged from the fourth through twelfth grade reading levels. The variety of reading levels in the science book chosen is consistent with typical science textbooks (Kroeger et al., 2009). During the intervention phases, students identified as competent in reading, based on a curriculum based assessment, read science passages with students possessing weaker reading skills.

The two teachers also modeled how to construct main ideas. After assembling their own main ideas, students documented them in their science journals during the PALScience Paragraph Shrinking activity (Kroeger et al., 2009). The teachers used a cloze procedure throughout the study to measure comprehension. Results indicated that PALScience improved student skills and overall performance on cloze activities. Although 70% of students responded on a survey that they did not like PALScience for various reasons such as, "You had to read," (Kroeger et al., 2009, p. 13) and "I didn't know my partner," (Kroeger et al., 2009, p. 13) 61% of students commented that PALScience helped them learn important things like, "How to read and comprehend paragraphs," (Kroeger et al., 2009, p. 13) and "To pay more attention to my reading" (Kroeger et al., 2009, p. 13).

PLAN. PLAN represents a study-reading program specifically used in middle school classrooms to improve comprehension of science texts. PLAN stands for Predict, Locate, Add, and Note. Students predict the content and structure of a text and assess its purpose by creating a diagram or probable map of the author's ideas as expressed in the chapter title, subtitles, highlighted words, and graphics (Caverly, Mandeville, & Nicholson, 1995). The map represents students' predictions of the importance of the chapter concepts and the ordered relationships among them. After students predict, they locate known and unknown information on the map by placing checkmarks next to familiar concepts and question marks by unfamiliar concepts. According to Caverly et al., (1995) as students read the chapter, they add words or short phrases to their map to explain the concepts marked with question marks and confirm and extend concepts with checkmarks. Finally, students take note of their new understanding by completing tasks such as summarizing the information, discussing it, and reconstructing a new map.

In one study researchers implemented PLAN in two science classes in a rural middle school. The study included a single-group pre-test/post-test design with one class of fifteen seventh-grade students and another of eighteen eighth grade students (Radcliff, Caverly, Peterson, & Emmons, 2004). Students were taught by a teacher who had completed a graduate course focusing on integrating reading strategies into content area teaching. The researchers modeled PLAN during one session of the graduate course and students in the course practiced the strategy in small groups. The teacher then implemented PLAN in two science classrooms during the fall term. In addition, the teacher met weekly, for a total of 15 hours, with researchers to discuss strategic textbook reading and its implementation into the middle school science classrooms (Radcliff et al., 2004).

Results revealed an increase in the percentage of propositions that reflected paraphrasing of content and higher order thinking on students' PLAN maps as PLAN was implemented in the classrooms. Results also revealed an increase in the reading strategy checklist scores from beginning to the end of the study. Interviews with the teacher and students in the study indicated that the PLAN program increased students' willingness and ability to learn from textbook reading (Radcliff et al., 2004).

In a second study the researchers used a nonequivalent-groups, pretest-posttest design to incorporate PLAN into one sixth grade classroom of 23 students and another sixth grade classroom of 27 students used as a control group. As in the previous research, the teacher in this study had participated in a graduate class focusing on how to implement PLAN in her classroom. The teacher met for a total of 15 hours with researchers to discuss the processes of strategic textbook reading and the challenges of implementing them into a middle school classroom (Radcliff, Caverly, Hand, & Franke, 2008). In this study, the teacher taught PLAN to one of her science classes while another class followed her traditional instruction without PLAN (Radcliff et al., 2008). Results indicated that the average score on the comprehension tests for students in the treatment group was significantly higher than the average score for students in the control group. This result occurred despite the fact that the pretests of groups did not differ significantly. A similar trend took place with the reading checklist scores (Radcliff et al., 2008). According to Radcliff et al. (2008) the reading checklist consisted of 10 yes or no questions regarding what strategies students used for reading a textbook chapter and for monitoring comprehension. Reading checklist scores improved significantly between pre and posttest scores for the treatment group but not for the control group. Pre test scores were similar for both groups but posttest scores were higher for the treatment group.

Inquiry based curriculum with reading**.** Much of the recent scholarship on science education has emphasized the importance of inquiry as well as reading in the development of science literacy (Fang & Wei, 2010). An inquiry- based curriculum recognizes science as a process for producing knowledge depending on careful observations and grounded interpretations. It also focuses on the development of skills in acquiring science knowledge, using high-level reasoning, applying existing understanding of scientific ideas, and communicating scientific information. Recently, science educators have expanded their concept of science literacy to include general reading ability because, without the ability to read, students are limited in the depth and breadth of scientific knowledge they can attain (Fang & Wei, 2010).

Fang and Wei (2010) examined the impact of an inquiry-based science curriculum that infused explicit reading strategy instruction in the science literacy development of middle school students. The study took place in ten regular sixth-grade science classes, with two teachers teaching five classes a piece. In three classes per teacher, explicit instruction of reading strategies was taught for an average of 15-20 minutes per week and students had access to a home reading program that encouraged them to read one quality science trade book per week. Each teacher's remaining two classes were used as control groups, provided with an inquiry based curriculum similar to the experimental groups, but without reading instruction (Fang & Wei, 2010). During the 15-20 minutes of reading strategy instruction, the researchers co-taught with teachers using an explain-model-guide-apply (EMGA) instructional program that maintained a review of the previous week's strategy, an explanation of the target strategy for the week, the teacher's modeling of the use of the strategy, and brief guided and independent practices of applying the strategy. Students then checked out a science trade book published by organizations such as the National Science Teachers Association in cooperation with the Children's Book Council, and were reminded daily to use the target strategy when reading the trade book as well as their science textbook (Fang & Wei, 2010). The students also participated in a review session at the end of the fall semester, in which they selected a strategy that was previously taught and used it with a text excerpt from a science trade book. They also participated in a second review session at the end of the spring semester when they got to comment on their favorite strategies.

Results revealed that not only did students in the inquiry-based science with reading curriculum demonstrate more knowledge about science content on the curriculum-referenced science posttest scores, but they also scored significantly higher on the Gates-MacGinitie Reading posttest. Thus, students' general knowledge of scientific content as well as their general reading ability improved based on the inquiry-based science plus reading curriculum (Fang & Wei, 2010).

Rogevich and Perin (2008) found a very effective program to improve science comprehension in students with behavioral disorders (BD) and/or attention deficit hyperactive disorder (ADHD). The program, called Thinking before reading, While reading, and After reading (TWA), was used in conjunction with written summarization (TWA-WS) at a residential treatment facility for adolescents adjudicated as juvenile delinquent by local courts (Rogevich & Perin, 2008). The study took place at the school on the campus of the residential treatment facility and 63 boys ranging from ages 13-16 participated. Thirty-two of the participants had been diagnosed with BD while 31 had been diagnosed with BD plus ADHD. The mean reading test score on the Gates-MacGinitie Reading Test indicated that students read on approximately a fifth grade level (Rogevich & Perin, 2008).

The investigators used a matched comparison design with one treatment group consisting of students with BD who received intervention, another treatment group consisting of students with BD plus ADHD who received intervention as well as two control groups. Control groups consisted of students with BD or BD and ADHD who received traditional literacy practice but not the experimental intervention.

The researchers conducted the study in a total of eight sessions, with three assessment sessions and five instructional sessions. In the TWA strategy with science passages, small groups of learners identified the author's purpose in a written text, determined what they already knew about the topic, set a reading goal, focused on their reading speed, linked their background knowledge to information from the text that was new to them, reread parts of the selection, identified the main idea, orally summarized the information in the text, and reflected on what they had learned (Rogevich & Perin, 2008). Students also completed the WS part of the strategy through pretest, posttest, near transfer, far transfer, and maintenance summarization tasks. Results indicated that TWA-WS was effective on all measures, as compared with practice with the same text without TWA-WS (Rogevich & Perin, 2008).

Implications for Research

The literature on iSTART provides very encouraging results about the use of the program to improve reading comprehension of adolescents and college students. iSTART benefits most students who participate in training, but it benefits them in different ways. All three studies we reviewed demonstrated the ability of iSTART to help skilled readers generate more bridging inferences and less skilled readers to gain a basic understanding of text-based information. Thus, iSTART appears to help students based on their zone of proximal development (McNamara et al., 2006). Future research on iSTART might include how the program could be adapted to the specific needs of less skilled and more skilled readers (Magliano et al., 2005). For example, future versions of iSTART could provide less skilled and low strategy knowledge students with more training in lower level strategies, and more positive feedback for strategies such as paraphrasing. On the other hand, future versions of iSTART could push more skilled readers to go beyond the text and use strategies such as elaboration to create coherence (McNamara et al., 2006).

Literature examining the benefits of peer-assisted learning programs suggests another possibility for improving middle and high school students' comprehension of science texts. Research indicates that PALS improves reading comprehension of struggling adolescent readers, especially those within special education classrooms. Implementation of a modified version of PALS, PALScience, resulted in an improvement in readers' comprehension of science texts. One very positive aspect of PALS is that teachers actually conduct the studies in their classrooms and not in a laboratory setting. Also, PALS can be adapted to use in a variety of grade levels. Researchers have implemented K-PALS for kindergarten students, First grade PALS for first grade students, as well as High School PALS. While studies involving PALS show an increase in reading comprehension, the literature also reveals areas of limitation. The small sample sizes of previous studies and the use of PALS primarily in special education classes prevent results from being as generalizable as desired. Finally, very little research about the use of PALS for expository texts is available.

The study-reading program PLAN represents another approach that shows promise in improving secondary students' comprehension of science texts. Research on PLAN indicates that students use more reading strategies when using PLAN and design PLAN maps demonstrating higher ordered thinking and better paraphrasing of material as time progresses. Radcliff et al. (2008) also showed an increase in reading comprehension tests for middle school students after participating in PLAN. However, research on PLAN is limited. Future research should include its use in high school science classrooms since previous studies have been conducted in middle schools.

The results of Fang and Wei's (2010) study contributed to previous research on reading comprehension of science texts because they conducted their study in a middle school general education science classroom instead of a reading classroom or laboratory. The authors also required that students practice the reading strategies they learned in class outside of the classroom with science textbooks and science trade books. Limitations of this study include the small amount of time spent in class per week teaching reading strategies. More class time may elicit greater reading comprehension and science knowledge. Future research with the inquiry-based science plus reading curriculum should consider more time spent each week in class on explicit reading strategies and greater implementation of the curriculum by science teachers trained in the program.

Limitations of the Rogevich and Perin (2008) study include a small sample size consisting of students with specific diagnoses. Future research should include a larger population of students, more than one teacher, and the inclusion of female students with BD and/or ADHD. Investigators should also examine whether TWA-WS would work well with students who do not possess BD and/or ADHD, to see whether the strategy is effective with a general education secondary population.

Conclusion

According to the NAEP, many secondary students lack the ability to master science concepts and research is clear that adolescents struggle to comprehend science textbooks. An inability to use appropriate reading strategies accounts for one reason why many adolescents cannot comprehend science textbooks. Literature suggests that improving the reading strategies of readers will improve their overall comprehension of science text information. Research provides evidence that students who receive reading strategy instruction show improvement in reading comprehension of science texts.

The purpose of this article was to review literature on reading programs that improve secondary students' comprehension of science texts. The reading program iSTART, PALScience, PLAN, TWA-WS, and an inquiry-based science curriculum plus reading were described in the article. Results from studies conducted using each of the programs were discussed and provide a great deal of support for use of the programs in improving adolescents' comprehension of science texts.

• • •

Brandi E. Johnson is a science teacher at Benjamin Elijah Mays High School in Atlanta, Georgia. She has a BA in Psychology from Spelman College and two master's degrees from Georgia State University, in Urban Teacher Leadership and Educational Psychology. She is currently a doctoral student at Clark Atlanta University in the department of Educational Leadership. Her research interests include science reading comprehension, single-gender classrooms and schools, culturally relevant pedagogy, and school culture.

Karen M. Zabrucky is a Professor of Educational Psychology at Georgia State University. Her academic background includes a Bachelor's degree in Psychology and Sociology, a Master's degree and Ph.D. in Experimental Psychology, and a Postdoctoral Fellowship in Applied Cognitive Psychology. Her research interests include text comprehension and memory, metacomprehension, and memory in the real world. She is on the editorial board of IEJEE and is delighted to serve as editor of IEJEE's special issue on reading comprehension.

References

Baker, L. (1985). How do we know when we don't understand? Standards for evaluating comprehension. In D. L. Forrest, G. E., MacKinnnon, & T. G. Waller (Eds.), *Metacognition, cognition, and human performance* (pp 155-205). New York: Academic Press.

Best, R. M., Flyod, R. G., & McNamara, D. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology*, *29*, 137-164.

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

Calhoon, M. (2005). Effects of peer-mediated phonological skill and reading comprehension program on reading skill acquisition for middle school students with reading disabilities. *Journal of Learning Disabilities*, *38*, 424-433.

Carnine, L., & Carnine, D. (2004). The interaction of reading skills and science content knowledge when teaching struggling secondary students. *Reading & Writing Quarterly, 20,* 203-218.

Caverly, D., Mandeville, T., & Nicholson, S. (1995). PLAN: A study-reading strategy for informational text. *Journal of Adolescent & Adult Literacy*, *39*, 190-199.

Chambliss, M. & Calfee, R. (1989). Designing science textbooks to enhance student understanding. *Educational Psychologist, 24,* 307-322.

Cook, L. & Mayer, R. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology, 80,* 448-456.

Fang, Z., & Wei, Y. (2010). Improving middle school students' science literacy through reading infusion. *The Journal of Educational Research*, *103*, 262-273.

Graesser, A., McNamara, D., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through point & query, autotutor, and iSTART. *Educational Psychologist, 40*, 225-234.

Graesser, A., Jeon, M., Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes*, *45*, 298-322.

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

Kroeger, S., Burton, C., & Preston, C. (2009). Integrating evidence-based practices in Middle science reading. *TEACHING Exceptional Children*, *41*, 6-15.

Magliano, J., Todaro, S., Millis, K., Wiemer-Hastings, K., Kim, H., & McNamara, D. (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.

McMaster, K., Fuchs, D., & Fuchs, L. (2006). Research on peer-assisted learning strategies: The promise and limitations of peer-mediated instruction. *Reading & Writing Quarterly*, *22*, 5-25.

McNamara, D., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Educational Computing Research, 31*, 147-171.

National Center for Education Statistics. (2010, March 12) *The Nation's Report Card: The NAEP Reading Achievement Levels by Grade*. Retrieved June 25, 2010, from http://nces.ed.gov/nationsreportcard/reading/achieveall.asp

National Center for Education Statistics. (2002, June 19) *The Nation's Report Card: The NAEP Science Achievement Levels*. Retrieved June 25, 2010, from http://nces.ed.gov/nationsreportcard/science/achieveall.asp

Radcliff, R., Caverly, D., Hand, J., & Franke, D. (2008). Improving reading in a middle school science classroom. *Journal of Adolescent & Adult Literacy, 51*, 398-408.

Radcliff, R., Caverly, D., Peterson, C., & Emmons, M. (2004). Improving textbook reading in a middle school science classroom. *Reading Improvement*, *41*, 145-156.

Rogevich, M., & Perin, D. (2008). Effects on science summarization of a reading comprehension intervention for adolescents with behavior and attention disorders. *Exceptional Children*, *74*, 135-154.

Zimmerman, C., Gerson, S., Monroe, A., & Kearney, A. M. (2007). Physics is harder than psychology (or is it?): Developmental differences in calibration of domain-specific texts. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the Twenty-ninth Annual Cognitive Science Society* (pp. 1683 – 1688). Austin, TX: Cognitive Science Society. http://www.cogsci.rpi.edu/ csjarchive/ proceedings /2007/docs/p1683.pdf

www.iejee.com

This page is intentionally left blank

International Electronic Journal of Elementary Education, 2011, 4(1), 33-46.

Translating advances in reading comprehension research to educational practice

Danielle S. McNAMARA*
Arizona State University, United States

Panayiota KENDEOU
Neapolis University Pafos, Cyprus

Abstract

The authors review five major findings in reading comprehension and their implications for educational practice. First, research suggests that comprehension skills are separable from decoding processes and important at early ages, suggesting that comprehension skills should be targeted early, even before the child learns to read. Second, there is an important distinction between reading processes and products, as well as their causal relationship: processes lead to certain products. Hence, instructional approaches and strategies focusing on processes are needed to improve students' reading performance (i.e., product). Third, inferences are a crucial component of skilled comprehension. Hence, children need scaffolding and remediation to learn to generate inferences, even when they know little about the text topic. Fourth, comprehension depends on a complex interaction between the reader, the characteristics of the text, and the instructional task, highlighting the need for careful selection of instructional materials for individual students and specific groups of students. Finally, educators may benefit from heightened awareness of the limitations and inadequacies of standardized reading comprehension assessments, as well as the multidimensionality of comprehension to better understand their students' particular strengths and weaknesses.

Keywords: Reading comprehension research, education practice, instructional implications

Introduction

Translating Advances in Reading Comprehension Research to Educational Practice

An extensive and impressive knowledge base has been established in the area of reading comprehension (for reviews, see McNamara & Magliano, 2009; RAND Reading Study Group, 2002). In the present paper, our aim is to discuss important findings in reading

* Danielle S. McNamara, Department of Psychology Senior Research Scientist, Learning Sciences Institute Arizona State University, P.O. Box 872111,Tempe, Arizona 85287-2111, United States.
E-mail: Danielle.McNamara@asu.edu

ISSN:1307-9298

www.iejee.com

comprehension research and particularly those findings that we deemed to have the potential to impact educational practice but have yet to be fully utilized. These findings are:

- Dissociations between decoding and comprehension skills
- Distinctions between the process and the product of reading comprehension
- The importance of prior knowledge to inferencing
- Interdependencies among reader and text characteristics
- Inadequacies of commonly used reading comprehension assessments

In the first section of the paper, we briefly summarize the findings for each of these advances in reading comprehension research. In the second section, we translate these advances into concrete recommendations for educational practice. In the final section, we discuss future directions for reading comprehension research.

What do we know about reading comprehension?
Comprehension versus Decoding

A prominent advance in reading comprehension research concerns the relation between decoding and comprehension skills. Many researchers have focused on the initial stages of reading acquisition, highlighting the importance of decoding skills - skills that support reading, such as phonological awareness, letter and word identification (for reviews, Snowling & Hulme, 2005; Storch & Whitehurst, 2002), and comprehension skills - skills that support oral language comprehension, such as receptive vocabulary and listening comprehension (e.g., Kendeou, van den Broek, White, & Lynch, 2009; Lonigan, Burgess, & Anthony, 2000).

The relation between these two sets of skills is expressed most succinctly within the Simple View of Reading (Gough, Hoover, Petersen, 1996; Gough & Tunmer, 1986). In the SVR, reading comprehension is described as the product of a reader's word decoding and listening comprehension skills, with the central tenet that both decoding and comprehension are necessary for reading comprehension. Notably, advocates of the SVR model do not discount other potential contributors to the reading process, but rather propose that decoding and comprehension are the core competencies (Kendeou, Savage, & van den Broek, 2009).

In principle, if decoding and comprehension are separate dimensions of reading comprehension, then each should depend on different underlying skills and abilities and children can perform differentially on these two sets of skills. Indeed, research on poor readers has identified children with good decoding but poor comprehension skills (Cain, Oakhill, & Lemmon, 2005; Nation, 2005; Stothard & Hulme, 1992) as well as children with poor decoding but good comprehension skills (Adlof, Catts, & Little, 2006; Spooner, Baddeley, & Gathercole, 2004).

In addition, several studies on the development of decoding and comprehension skills have suggested their dissociation (Cutting & Scarborough, 2006; Kendeou, van den Broek, White, & Lynch, 2007; Kendeou et al., 2005, 2009; Muter, Hulme, Snowling, & Stevenson, 2004). There is convincing evidence that these sets of skills are separate and relatively unrelated from preschool to early elementary school (Catts, Fey, Zhang, & Tomblin, 1999; Cain, Oakhill, & Bryant, 2004; Kendeou et al., 2005, 2007, 2009; Savage, 2006), and both set of skills significantly, and independently contribute to reading comprehension performance in early elementary school (Kendeou et al., 2009; Storch & Whitehurst, 2002).

Process versus Product of Reading Comprehension

Reading comprehension research has produced detailed and valuable information regarding the development of reading comprehension skills and the factors that influence and are influenced by these skills (McNamara & Magliano, 2009). Although there are many definitions of what constitutes successful reading comprehension, a common component of most definitions is that it involves the construction of a coherent mental representation of the text in readers' memory. This mental representation is the product of reading comprehension. Its construction, however, is the *process* of comprehension and occurs moment-by-moment as the individuals read. Distinguishing between products and processes is important because the two are causally related: Reading processes lead to reading products (Kintsch, 1988; Trabasso & Suh, 1993). Most important, failures in particular processes can lead to comprehension difficulties, and by consequence, low performance in terms of its products.

In reading comprehension research, a focus solely on products limits our ability to identify underlying mechanisms that may lead to changes in reading performance (Magliano, Millis, Ozuru, & McNamara, 2007). A focus solely on processes limits our ability to determine potential impacts of textual, reader, and task factors on reading performance. By considering both processes and products, researchers provide increased rigor in the investigation of various issues, as well as a deeper understanding of how to best facilitate reading comprehension. Importantly, by knowing at which points and why the process fails, we can design appropriate interventions and learning materials to prevent or remediate the problem.

Inferencing and Prior Knowledge

Another major advance in reading comprehension research concerns the central role of inference processes and the role of prior knowledge in these processes. Indeed, the ability to draw inferences is central to reading comprehension across the lifespan (Oakhill, Cain, & Bryant, 2003; Paris, Lindauer, & Cox, 1977; van den Broek, 1990), and there is direct evidence that it is not just a by-product of comprehension, but rather a plausible cause (Cain & Oakhill, 1999).

In the context of reading comprehension, *inferencing* is the process of connecting information within the text or within the text and one's knowledge base, and drawing a conclusion that is not explicitly stated in the text. One type of inference, called bridging inferences, connects current text information to information that was previously encountered in the text, such as connecting the current sentence to a previous sentence. Another type, associative inferences or elaborations, connects current text information to knowledge that is not in the text. For such knowledge-based inferences, readers bring knowledge that is related to the text to the focus of attention, and in doing so, construct connections between the text and prior knowledge (Cook, Limber, & O'Brien, 2001). Of course, a reader continuously draws upon knowledge with every word encountered in a text. Each word requires accessing memory to process its meaning. Prior knowledge of both the words in the text and related concepts are activated or *primed* (O'Brien & Myers, 1999). When comprehenders have more knowledge about the domain, or about the world, then their understanding of a text or discourse is likely to be richer and more coherent because more concepts that are not explicit in the text are available to the reader and become part of the reader's mental representation of the text.

Consequently, readers' prior knowledge directly influences readers' ability to generate inferences. Readers who have more knowledge about the topic of a text better understand the written material (Chiesi, Spilich, & Voss, 1979; Haenggi & Perfetti, 1994) and are better able to comprehend texts that require numerous inferences (McNamara, 2001; O'Reilly &

McNamara, 2002). These advantages partially arise because high knowledge readers answer comprehension questions based on prior knowledge, rather than information in the text and are more likely to generate inferences that connect new information in the text with prior knowledge.

Knowledge helps a reader to make inferences. In addition, the reader also needs to know how to make inferences. Behavioral studies of individual differences in comprehension indicate that skilled and less-skilled readers differ primarily in terms of inference processes such as solving anaphoric reference, selecting the meaning of homographs, processing garden-path sentences, and making appropriate inferences while reading (Long, Oppy, & Seely, 1994; Oakhill, 1984; Oakhill & Yuill, 1996; Singer & Ritchot, 1996; Whitney, Ritchie, & Clark, 1991; Yuill & Oakhill, 1988). Protocol analyses have further revealed that skilled readers are also more likely to generate inferences that repair conceptual gaps between clauses, sentences, and paragraphs (Magliano & Millis, 2003; Magliano, Wiemer-Hastings, Millis, Muñoz, & McNamara, 2002). In contrast, less skilled readers tend to ignore conceptual gaps in text while reading and often fail to make the inferences necessary to fill in the gaps (Garnham, Oakhill, & Johnson-Laird, 1982; Oakhill, Yuill, & Donaldson, 1990).

Reader-Text Interactions

There are innumerable factors affecting reading comprehension, such as reader characteristics, text properties, and the instructional context in which reading takes place (Dixon & Bortolussi, 1996; RAND Reading Study Group, 2002; Stanovich & Cunningham, 1993; van den Broek & Kremer, 1999). Although these factors have often been studied in isolation, a consideration of their interactions and interdependencies provides crucial information about the comprehension process (Kintsch, 1998; RAND Reading Study Group, 2002).

Among text characteristics, text cohesion is an important aspect of text that influences reading comprehension processes. Cohesion arises from a variety of sources, including explicit referential overlap and causal relationships (Givón, 1995; Graesser, McNamara, & Louwerse, 2003). Referential cohesion, for example, refers to the degree to which there is overlap or repetition of words or concepts across sentences, paragraphs, or the entire text. McNamara, Kintsch, Songer, and Kintsch (1996) found that *the effects of text cohesion and reader prior knowledge interact* (O'Reilly & McNamara, 2007a; Ozuru, Dempsey, & McNamara, 2009). These studies show that low-knowledge readers benefit from added textual cohesion because they lack the necessary knowledge to generate inferences. By contrast, high-knowledge readers (i.e., who do not generate strategic inferences; O'Reilly & McNamara, 2007a) benefit from cohesion gaps in the text because they are induced by the gaps to generate inferences.

Other investigators have similarly demonstrated that comprehension is enhanced when readers are induced by the text to generate inferences and these inferences are successful (Einstein, McDaniel, Owen, & Cote, 1990; Mannes & Kintsch, 1987; O'Brien & Myers, 1985; Rauenbusch & Bereiter, 1991). The theoretical explanation for these findings rests on the assumption that comprehension is largely determined by the coherence of the reader's mental representation of the text, and this is a function of both the ease of processing the text and the inferences generated by the reader. As illustrated in Figure 1, McNamara and Magliano (2009) proposed that reading comprehension will tend to be best when the ease of processing is high and the reader is strategic (quadrant A), and worse when the ease of processing is low and the reader is not strategic (quadrant D). Comprehension will tend to be more superficial and thematic in quadrant B, and will tend to be limited more to a textbase level understanding (i.e., a representation that primarily reflects the explicit content presented in the text) in quadrant C.

Of course, these comprehension outcomes are a function of numerous other interdependent factors that influence reading. For example, the ease of processing a text can depend on the familiarity of the words, the complexity of the domain, text readability, text cohesion, text domain or genre, and a many other factors, some of which depend on each other producing complex interactions. Similarly, the likelihood that a reader will engage in strategic comprehension processes can depend on reading skill, comprehension skill, motivation, metacognitive awareness, domain knowledge, reading strategy knowledge, goals, and tasks, which in turn can interact, not only with one another but with characteristics of the text. Thus, the causes vary for a text being more or less facile to understand and for a comprehender to be more or less strategic. Nonetheless, narrowing the focus on these two overarching factors (i.e., text ease, strategic processing) provides a heuristic for better understanding the scope of comprehension outcomes across studies and situations.

Assessment of Reading Comprehension

The assessment of reading comprehension has been one of the most controversial issues in the field (Keenan, in press; Kendeou & Papadopoulos, in press) for several reasons. First, the assessment typically focuses on the product of reading and provides little or no information with respect to the actual processes (Magliano et al., 2007; Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007). Second, it is the complex interaction of many factors, such as types of texts and response formats that influences students' performance (Ozuru, Rowe, O'Reilly, & McNamara, 2008; Paris, 2007). Third, current assessments confound comprehension with vocabulary, prior knowledge, word decoding, and other reader abilities involved in comprehension. Finally, well-known tests of reading comprehension have been criticized for lacking content and concurrent validity (Keenan & Betjemann, 2006) and for differential dependencies on decoding and comprehension skills (Keenan, Betjemann, & Olson, 2008). These controversial issues highlight that the measurement of the construct is not a trivial task.

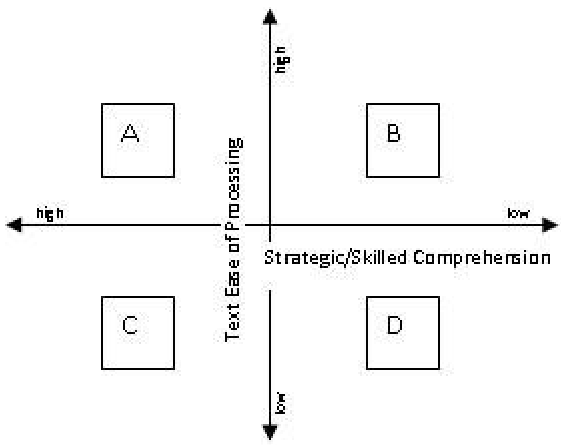


Figure 1. *Four quadrants crossing text ease and reader abilities (reprinted from McNamara & Magliano, 2009).*

On the one hand, the assessment of reading comprehension in research studies has often been theory-based and guided by comprehension models such as Kintsch and van dijk's (1978; Kintsch, 1998), which propose that there are multidimensional levels of understanding that emerge during the comprehension process, including surface, textbase, and situation model levels (McNamara & Magliano, 2009). By consequence, in order to obtain a full picture of a reader's understanding, assessment needs to consider the full range of these potential levels of comprehension. Although comprehension is assumed to be one interconnected mental representation, different types of assessments pull out different levels of understanding. For example, multiple-choice questions that target a restricted range of text (e.g., a single sentence or consecutive sentences) tend to provide an indication of the degree to which the reader can recognize the information from the text, usually at superficial level (Keenan et al., 2008). Cloze tasks that require readers to fill in missing words in text (e.g., Woodcock-Johnson Test) assess comprehension only within sentences based on word associations (Shanahan, Kamil, & Tobin, 1982). At the other extreme, asking the reader to use the information in the text to solve a problem taps into the reader's deeper, situation model understanding of the text (McNamara et al., 1996).

Translating what we know to educational practice

We know a good deal more about comprehension than what could be presented in this paper (RAND Reading Study Group, 2002). This summary has presented a subset of what may be considered the most important findings regarding comprehension, and particularly those findings that may potentially have the most impact on educational practice. These findings can be translated into concrete recommendations for educational practice geared towards improving students' ability to understand and learn from text:

- Give early focus to comprehension skills
- Design interventions that influence the actual comprehension process
- Teach students to make inferences, even without prior knowledge
- Consider during instruction not only the reader and the text, but also their respective interaction
- Interpret student reading performance by considering the test used

Give early focus to comprehension skills

The literature on the relation between decoding and comprehension skills, and the contribution of those skills to later reading comprehension, highlights the importance of developing these skills in young children. These findings imply that both decoding and comprehension skills should be targeted well before the child can read fluently (Kendeou et al., 2007, 2009). Although much attention in preschool language programs has been devoted to basic language skills that support decoding, attention to comprehension skills is equally important and needs to be included in such programs.

Activities situated around television viewing or aural listening may provide the opportunity for developing comprehension skills that could later transfer to reading (Kendeou et al., 2005, 2007, 2009; van den Broek, Rapp, & Kendeou, 2005). The use of television or aural stories offers several advantages over the use of printed text alone for young children because they are highly motivating and can easily be used with a large group of children in school as well as non-school (e.g., home) settings. The use of these non-written media provides a unique opportunity for children to be taught comprehension strategies that are not completely dependent upon verbal skills. Also, well-known reading comprehension instructional programs such as Reciprocal Teaching (Palincsar & Brown, 1984) and Peer Assisted Learning Strategies (PALS; Fuchs, Fuchs, Mathes, & Simmons, 1997) can be modified and used to foster comprehension development even at preschool age.

Design interventions that influence the actual comprehension process

The review on the distinction between process and product of reading comprehension highlights that the product of reading is directly influenced by the processes that take place during reading (Kintsch, 1998; Trabasso & Suh, 1993; van den Broek et al., 2005). This finding implies that for teachers to be able to affect the product of reading (e.g., increase reading performance), they need to implement appropriate interventions and instructional approaches to affect the actual processes *during* reading.

Indeed, a wide array of strategy interventions that have been shown to be effective in elementary school instruction share this characteristic: *they influence the actual processes while reading unfolds* (McNamara, 2007; Pressley, 1998, 2000). For instance, some interventions emphasize the importance of asking questions during reading (King, 2007). Other methods have used group activities to help students learn to make connections while listening or reading (Palincsar & Brown, 1984; Fuchs & Fuchs, 2007; Yuill, 2007). Numerous activities can be used (in reading and non-reading contexts) to help students learn how to make connections and, as a result, construct better mental representations of the texts, including: reading and thinking-aloud activities, question asking, and paragraph summarization (Kendeou et al., 2007).

Teach students to make inferences, even without prior knowledge

The review on the central role of inferences in reading comprehension highlights the need for teachers to scaffold these skills to students of all ages, so they learn to generate inferences that connect ideas in the text, bring in prior knowledge, and construct connections using general knowledge, despite a lack of sufficient domain specific knowledge. iSTART is a computer-based technology that focuses on providing high school and college students with instruction and practice using reading comprehension strategies that compensate for knowledge deficits (McNamara, Boonthum, Levinstein, & Millis, 2007). Its design is based on a classroom intervention called Self-Explanation Reading Training (SERT; McNamara, 2004) that combines self-explanation (Chi et al., 1994) with reading comprehension strategies (Bereiter & Bird; 1985; Palincsar & Brown, 1984). Students learn to self-explain using five reading strategies: *monitoring comprehension*, *paraphrasing*, making *bridging inferences* between the current sentence and prior text, making *predictions*, and *elaborating* the text with links to what the reader already knows. Studies evaluating iSTART's impact indicate that both strategy use and comprehension are enhanced (McNamara et al., 2007). However, the locus of the effect depends on the student's prior abilities. Low knowledge, less skilled students benefit most at the textbase level of comprehension, whereas high-knowledge, more skilled students show the largest gains on deep level questions (Magliano et al.,2007; McNamara et al., 2006; O'Reilly, Best, & McNamara, 2004). Thus, students benefit from reading strategy training at their zone of proximal development.

When teachers employ interventions that target more active or strategic use of knowledge, students' reading skill and comprehension can be dramatically improved (Bereiter & Bird, 1985; Chi, de Leeuw, Chiu, & LaVancher, 1994; Cornoldi & Oakhill, 1996; Dewitz, Carr, & Patberg, 1987; Hansen & Pearson; 1983; Kucan & Beck, 1997; McNamara, 2004; Palincsar & Brown, 1984; Paris, Cross, & Lipson, 1984; Yuill & Oakhill, 1988). In essence, with strategy interventions (e.g., SERT) teachers assist students to learn *how* and *when* to make inferences. Such training could be achieved in a classroom, for example, by asking students in pairs to take turns self-explaining a portion of the textbook or having students explain as a class while maintaining a continuum. If students use and practice the strategies, the potential benefit to their performance is substantial.

Consider during instruction not only the reader and the text, but also their respective interaction

The review on the interactions and interdependencies between readers and texts highlights the complexity of reading comprehension and the need for careful selection of instructional materials for individual students and specific groups of students. Students may benefit simply from educators' heightened awareness of the complexity of factors that can influence comprehension as well as a better understanding of the importance of choosing the right texts for the right students. Optimally, students need to be provided with texts that they can understand.

Considering Figure 1, if educators choose relatively easy-to-read texts and their students are highly motivated and strategic readers (quadrant A), then the students can be expected to develop a relatively deep understanding (Graesser, Singer, & Trabasso, 1994; Guthrie & Alao, 1997). If, on the other hand, educators choose relatively easy texts and their students are less-skilled, or less motivated to engage in strategic processes (quadrant B), then students can be expected to develop a more superficial, or textbase level of understanding (McKoon & Ratcliff, 1992). Likewise, strategic or skilled students can be expected to construct a relatively coherent textbase when they face a challenging, knowledge-demanding text (quadrant C; McNamara, 2004; O'Reilly & McNamara, 2007a, 2007b). By contrast, given a text with relatively familiar material containing numerous conceptual gaps (i.e., the reader is high in knowledge), students can be expected to develop a coherent situation model because students with sufficient knowledge will be able to generate the gap-filling inferences (McNamara et al., 1996). The worst levels of comprehension can be expected for less-skilled or unmotivated readers who encounter overly challenging texts. Students without sufficient prior knowledge, and who also lack sufficient reading comprehension skills, can be expected to understand little from text (O'Reilly & McNamara, 2007a).

Interpret student reading performance by considering the test used

Our review on the assessment of reading comprehension has highlighted that students' likelihood of success in reading and comprehension depends on the material that is read and the task that is completed during or after reading this material (Fletcher, 2006). Research indicating that common tests of comprehension do not tap into the same array of language and cognitive processes suggests that performance on reading comprehension tests may be influenced to different degrees by particular skills and different processes during comprehension.

These findings imply that different measures of reading comprehension can yield useful information for the educator *only* if we know the exact set of skills on which students' performance on the specific measure depends (Kendeou & Papadopoulos, 2008, in press). In the absence of such information, these measures provide only a basic indication of how well a student understands text and offer very little information about why some students may struggle while others succeed.

A departure point is an awareness that various assessments will provide more or less information about a student's abilities, and different information depending on the particular assessment. Nonetheless, one constraint faced by educators is that common and available measures of students' comprehension abilities generally provide a single score, under somewhat artificial, unmotivating circumstances. Educators need access to assessments that are indicative of the students' ability to draw inferences and build coherent mental representations of text. They also need access to comprehension assessment techniques that are likely to reflect a student's deep understanding of material.

For example, oral retelling after reading a text can be reflective of a student's deeper level understanding (Kendeou et al., 2009). The underlying assumption of retelling is that when readers retell a story, they draw on their mental representation of the text read. This mental representation is the product of inferencing during which readers interconnect the events in the text with their prior knowledge, which primarily depends on the reader's situation model level of understanding. There are numerous other approaches to comprehension assessment that will tap into deeper levels of comprehension, such as summarization, self-explanation, challenging comprehension questions, problem solving questions, and essay writing.

Conclusions

Our understanding of comprehension has matured based on decades of research on text and discourse processing. An ongoing challenge is the translation of those findings into practice. Our list summarizes what we consider to be a few of the most important findings and their corresponding translations to educational practice. Of course we know much more about comprehension than summarized in this brief review (McNamara & Magliano, 2009; Perfetti, 1985; Pressley, 1998, 2000; RAND Reading Study Group, 2002).

In addition, research on reading comprehension can inform educational practice, which in turn, should feed back to basic research in reading comprehension. For example, *effective reading comprehension instruction in mixed-ability classrooms* has a large impact on standardized measures of reading comprehension. Reading comprehension research needs to be informed directly as to the factors that lead to a broad impact on students' reading comprehension. This will necessitate a comprehensive examination of classrooms as complex systems, taking into account the teacher, the students, the approaches, the materials, and their respective interactions. A second example is *the relation between reading comprehension and writing*. In educational settings, the two are closely linked and taught building upon one another. Research in these two areas, however, has been relatively unconnected. Both researchers and educators would benefit from increased cross-talk among these areas and a consideration of how each influences and is being influenced by the other.

Perhaps one of the largest gaps in our understanding of comprehension processes regards how students process and understand information in multimedia environments (Mayer, 2001; McNamara & Magliano, 2009; van den Broek, Kendeou, & White, 2009). This dearth in research ranges from the students' processing of texts and pictures, to understanding material on the web, to the integration of information from various sources and mediums (e.g., text, video, discourse, pictures). These are areas we expect to garner increased attention by researchers in the near future.

• • •

Danielle S. McNamara is a Professor in the Psychology Department and Senior Scientist in the Learning Sciences Institute at Arizona State University. Her academic background includes a Linguistics B.A. (1982), a Clinical Psychology M.S. (1989), and a Cognitive Psychology Ph.D. (1992). Her research involves the study of learning and cognitive sciences and the development of educational technologies (Coh-Metrix, iSTART, Writing Pal; see soletlab.com). One focus of her work is on developing methods to improve success for struggling high school students. She has served on numerous funding panels, governing boards, editorial boards, and as Associate Editor for three journals.

Panayiota Kendeou is an Assistant Professor of Educational Psychology at Neapolis University Pafos, Cyprus. Her current research focuses on the cognitive processes that support memory and learning in the context of reading comprehension. She is currently an Associate Editor for Journal of Research in Reading and on the editorial boards of Scientific Studies of Reading, Contemporary Educational Psychology, Learning and Instruction, and Reading Psychology.

References

Adlof, S.M., Catts, H.W., & Little, T. (2006). Should the simple view of reading include a fluency component? *Reading and Writing: An Interdisciplinary Journal, 19,* 933-958.

Bereiter, C. & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction, 2,* 131-156.

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

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., (2005). The relation between children's reading comprehension level and their comprehension of idioms. *Journal of Experimental Child Psychology*, *90*, 65-87.

Catts, H.W., Fey, M.E., Zhang, X., & Tomblin, J.B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3,* 331-361.

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

Chiesi, H.L., Spilich, G.J., & Voss, J.F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*, 257-273.

Cook, A.E., Limber, J.E., & O'Brien, E.J. (2001). Situation-based context and the availability of predictive inferences. *Journal of Memory & Language, 44,* 220-234.

Cornoldi, C. & Oakhill, J.V. (Eds.). (1996). Reading comprehension difficulties: Processes and remediation. Mahwah, NJ: Lawrence Erlbaum Inc.

Cutting, L.E. & Scarborough, H.S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 227-299.

Dewitz, P., Carr, E., & Patberg, J. (1987). Effects of interference training on comprehension and comprehension monitoring. *Reading Research Quarterly, 22*, 99-121.

Dixon, P. & Bortolussi, M. (1996). Literary communication: Effects of reader-narrator cooperation. *Poetics, 23*, 405-430.

Einstein, G.O., McDaniel, M.A., Owen, P.D., & Coté, N.C. (1990). Encoding and recall of texts: The importance of material appropriate processing. *Journal of Memory and Language, 29*, 566-581.

Fletcher, J.M. (2006). Measuring reading comprehension. *Scientific Studies of Reading, 10*, 323-330.

Fuchs, L.S., & Fuchs, D. (2007). Instruction on mathematical problem solving. In D. Berch & M. Mazzacco (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 397-414). Baltimore, MD: Brookes.

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.

Garnham, A., Oakhill, J.V., & Johnson-Laird, P.N. (1982). Referential continuity and the coherence of discourse. *Cognition, 11*, 29-46.

Givón, T. (1995). Coherence in text vs. coherence in mind. In M.A. Gernsbacher & T.

Givón (Eds.), *Coherence in spontaneous text* (pp. 59–115). Amsterdam: John Benjamin.

Gough, P.B., Hoover, W.A., & Peterson, C.L. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 1-13). Mahwah, NJ: Erlbaum.

Gough, P.B. & Tunmer, W.E. (1986). Decoding, reading, and reading disabilities. *Remedial and Special Education, 7*, 6-10.

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. Sweet & C.E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82-98). New York, NY: Guilford Publications.

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

Guthrie, J.T. & Alao, S. (1997). Designing contexts to increase motivations for reading. *Educational Psychologist, 32*, 95-105.

Haenggi, D. & Perfetti, C.A. (1994). Processing components of college level reading comprehension. *Discourse Processes, 17*, 83-104.

Hansen, J. & Pearson, P.D. (1983). An instructional study: Improving the inferential comprehension of fourth grade good and poor readers. *Journal of Educational Psychology, 75*, 821-829.

Keenan, J. M. (in press). Measure for measure: Challenges in assessing reading comprehension. In J. Sabatini & E. Albro (Eds.), *Assessing reading in the 21st century: Aligning and applying advances in the reading and measurement sciences.*

Keenan, J.M. & Betjemann, R.S. (2006). Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading, 10*, 363-380.

Keenan, J.M., Betjemann, R.S., & Olson, R.K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*, 281-300.

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.

Kendeou, P. & Papadopoulos, T.C. (2008, November). Cognitive profiles of struggling readers: A theory-based approach to reading disabilities. Paper presented at the Annual Meeting of the Psychonomic Society, Chicago, IL.

Kendeou, P. & Papadopoulos, T.C. (in press). The use of CBM-Maze in Greek: A closer look at what it measures. In C. Espin, K. McMaster, & S. Rose (Eds.), *The progress monitoring paradigm: Research, policy, and practice.*

Kendeou, P., Savage, R., & van den Broek, P. (2009). Revisiting the simple view of reading. *British Journal of Educational Psychology, 79*, 353-370.

Kendeou, P., van den Broek, P., White, M.J., & Lynch, J.S. (2007). Preschool and early elementary comprehension: Skill development and strategy interventions. In D.S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies,* (pp. 27-45). Mahwah, NJ: Erlbaum.

Kendeou, P., van den Broek, P., White, M.J., & Lynch, J.S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*, 765-778.

King, A. (2007). Beyond literal comprehension: A strategy to promote deep understanding of text. In D.S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 267-290). Mahwah, NJ: Erlbaum.

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*. New York, NY: Cambridge University Press.

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

Kucan, L. & Beck, I.L. (1997). "Thinking aloud and reading comprehension research: Inquiry, instruction, and social Interaction. *Review of Educational Research, 67*, 271-299.

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.

Lonigan, C.J., Burgess, S.R., & Anthony, J.L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent variable longitudinal study. *Developmental Psychology, 36,* 596-613.

Magliano, J.P. & Millis, K.K. (2003). Assessing reading skill with a think-aloud procedure. *Cognition and Instruction, 21,* 251-283.Magliano, J.P., Millis, K.K.,

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, & Computers, 34*, 181-188.

Mannes, S.M. & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction, 4*, 91-115.

Mayer, R.E. (2001). *Multimedia Learning.* New York, NY: Cambridge.

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

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

McNamara, D.S. (2007). IIS: A marriage of computational linguistics, psychology, and educational technologies. In D. Wilson & G. Sutcliffe (Eds.), *Proceedings of the twentieth International Florida Artificial Intelligence Research Society Conference* (pp. 15-20). Menlo Park, California: The AAAI Press.

McNamara, D.S., Boonthum, C., Levinstein, I.B., & Millis, K. (2007). Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In T. Landauer, D.S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 227-241). Mahwah, NJ: Erlbaum.

McNamara, D.S., Kintsch, E., Songer, N.B., & 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. & Magliano, J.P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 297-384). New York, NY: Academic Press.

Muter, V., Hulme, C., Snowling, M.J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Developmental Psychology, 40*, 665-681.

Nation, K. (2005). Children's reading comprehension difficulties. In M.J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp 248-265). Oxford, UK: Blackwell Publishing.

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

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

Oahill, 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.* Mahwah, NJ: Erlbaum.

Oakhill, J., Yuill, N. & Donaldson, M. (1990). Understanding of "because" in skilled and less-skilled text comprehenders. *British Journal of Developmental Psychology, 8,* 401-410.

O'Brien, E.J. & Myers, J.L. (1985). When comprehension difficulty improves memory for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition 11,* 12-21.

O'Brien, E.J. & Myers, J.L. (1999). Text comprehension: A view from the bottom up. In S.R. Goldman, A.C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 35-54). Mahwah, NJ: Erlbaum.

O'Reilly, T. & McNamara, D.S. (2002). What's a science student to do? In W.D. Gray & C.D. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 726–731). Austin, TX: Cognitive Science Society.

O'Reilly, T., Best, R., & McNamara, D. S. (2004). Self-explanation reading training: Effect for low-knowledge readers. In K. Forbus, D. Gentner, & T. Regier (Eds.), Proceedings of the Twenty-

Sixth Annual Conference of the Cognitive Science Society (pp. 1053–1058). Mahwah, NJ: Erlbaum.

O'Reilly, T., & McNamara, D. S. (2007a). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes, 43*, 121-152.

O'Reilly, T. & McNamara, D.S. (2007b). 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, 44,* 161-196.

Ozuru, Y., Dempsey, K., & McNamara, D.S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction, 19,* 228-242.

Ozuru, Y., & McNamara, D.S. (2007). A multidimensional framework to evaluate reading assessment tools. In D.S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 107-136). Mahwah, NJ: Erlbaum.

Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D.S. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods, 40,* 1001-1015.

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

Paris, S.G. (2007). Assessment of reading comprehension. *Encyclopedia of language an literacy development* (pp. 1-8). London, ON: Canadian Language and Literacy Research Network.

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.

Paris, S.G., Lindauer, B.K., & Cox, G.L. (1977). The development of inferential comprehension. *Child Development, 47*, 1728-1733.

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

Pressley, M. (1998). *Reading instruction that works: The case for balanced teaching.* New York, NY: Guilford.

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: Erlbaum.

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

Rapp, D.N., van den Broek, P., McMaster, K.L., Kendeou, P., & Espin, C.A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading, 11*, 289-312.

Rauenbusch, F. & Bereiter, C. (1991). Making reading more difficult: A degraded text microworld for teaching reading comprehension strategies. *Cognition and Instruction, 8,* 181-206.

Savage, R. (2006). Reading comprehension is not always the product of nonsense word decoding and linguistic comprehension: Evidence from teenagers who are extremely poor readers. *Scientific Studies of Reading, 10*, 143-164.

Shanahan, R., Kamil, M.L., & Tobin, A.W. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly, 2*, 229-55.

Singer, M. & Ritchot, K.F.M. (1996). The role of working memory capacity and knowledge access in text inference processing. *Memory & Cognition, 24*, 733-743.

Spooner, A., Baddeley, A.D., & Gathercole, S.E. (2004). Can reading accuracy and comprehension be separated in the Neale Analysis of reading ability? *British Journal of Educational Psychology, 74*, 187-204.

Snowling, M.J. & Hulme, C. (2005) Learning to read with a language impairment. In M.J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 397-412). Oxford, UK: Blackwell Publishing.

Stanovich, K.E. & Cunningham, A.E. (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. *Journal of Educational Psychology, 85,* 211-229.

Storch, S.A. & Whitehurst, G.J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*, 934-947.

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.

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

van den Broek, P. (1990). The causal inference maker: Towards a process model of inference generation in text comprehension. In D.A. Balota, G.B. Flores d' Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 423-445). Hillsdale, NJ: Lawrence Erlbaum Associates.

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: Teacher's College Press.

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., Kendeou, P., & White, M. J. (2009). Cognitive processes during reading: Implications for the use of multimedia to foster reading comprehension. In A. G. Bus & S. B. Neuman (Eds.), *Multimedia and literacy development: Improving achievement for young learners* (pp. 57-73). New York: Rutledge.

Whitney, P., Ritchie, B., & Clark, M. (1991). Working memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes, 14,* 133–145.

Yuill, N. & Oakhill, J. (1988). Understanding of anaphoric relations in skilled and less skilled comprehenders. *British Journal of Psychology, 79,* 173–186.

International Electronic Journal of Elementary Education, 2011, 4(1), 47-65.

The role of verbal working memory in second language reading fluency and comprehension: A comparison of English and Korean

Hye K. PAE*
University of Cincinnati, United States

Rose A. SEVCIK
Georgia State University, United States

Abstract

This study examined the respective contribution of verbal working memory, which was operationalized as immediate digit and sentence recall, to bilingual children's reading fluency and comprehension in the first language (L1) and second language (L2). Fifty children from two international sites took part in this study: One group was English-Korean bilinguals in the U.S., while the other was Korean-English bilinguals in Korea. The manifestation of the prediction model varied across the learning contexts or learner groups. L1 forward and backward digit spans accounted for the significant variances in L2 reading fluency and comprehension for the English-speaking children in the U.S., whereas L1 forward digit span was more predictive of L2 reading fluency and comprehension than backward digit span and sentence recall for the Korean-speaking counterparts in Korea. The results were interpreted with respect to the orthographic depth, linguistic differences, and cognitive demands. Implications and future directions are discussed.

Keywords: Reading fluency and comprehension; English-Korean bilinguals; Korean-English bilinguals; verbal memory

Introduction

Since Baddeley and Hitch's publication (1974) on the construct of working memory, researchers have investigated not only the nature, structure, and function of working memory but also its relation to children's language and reading acquisition (Daneman & Carpenter, 1980; Gathercole & Baddeley, 1993; Windfuhr & Snowling, 2001). Working memory refers to the transitory storage capacity and operations that manipulate verbal or written input while processing incoming information and retrieving relevant phonological information from the long-term lexicon (Miller & Kupfermann, 2009). It works as a processing

* Hye K. Pae, School of Education, University of Cincinnati, Cincinnati, OH 45221, United States. Phone: 513-556-7112, E-mail: Hye.pae@uc.edu

ISSN:1307-9298

www.iejee.com

source for the active maintenance of task-relevant information while simultaneously processing the same or other information activated along with task operations (Swanson, Zheng, & Jerman, 2009). Under the working memory model (Baddeley, 1986), verbal short-term memory or the phonological loop captures a subset of working memory performance. This subsystem is usually measured using immediate serial recall tasks in which the examinee is verbally presented with the sequence of isolated digits or words (Gupta, 1996).

Learning relies on an individual's ability to conceptualize and categorize new information and to make associations with other information housed in mental storage. Given the importance of mental storage, working memory has received abundant attention over time. Limited auditory memory span is one of the sources of deficiencies in language processing because a reduced memory span can inhibit the efficiency of working memory, which is necessary for processing and comprehending extended verbal narrations (Gathercole & Baddeley, 1993; Miller & Kupfermann, 2009).

There have been conflicting views on the definitions of working memory and short-term memory. Some researchers (Colom, Flores-Mendoza, Quiroga, & Privado, 2005; Gathercole, 1998; Unsworth & Engle, 2007) have differentiated working memory from short-term memory, while others (McDougall, Hulme, Elllis, & Monk, 1994) have used the two terms interchangeably. The former's rationale is based on the claim that working memory relies on the central executive system with a heavy demand of information manipulation. In other words, short-term memory works to retrieve a sequence of items in the order in which information is stored without manipulations, while working memory requires recalling information that is transformed from its initial encoding in order to perform task-relevant operations through manipulations (Swanson, Zheng, & Jerman, 2009). The latter upholds a claim that working memory and short-term memory share some commonalities on the learner's part with regard to the speech-based phonological input memory bank and extra-transformational processes. Regardless of the position on the relationship between working memory and short-term memory, working memory and short-term memory are forms of transient memory. Since this study was not designed to test memory models, verbal memory was operationalized as transitory immediate digit and sentence recall or memory in this paper. The purpose of this study was to investigate the manifestation of cognitive demands necessary for first language (L1) and second language (L2) reading.

Relationships Between Phonological and Verbal Memory and Reading

Given the way that phonological coding contributes to the retention of information in working memory and vice versa, studies have demonstrated a significant association between verbal memory span and word reading (Jeffries & Everatt, 2004; Johnston, Rugg, & Scott, 1987; Jorm, Share, Maclean, & Matthews, 1984). Deficits in verbal working memory and slow or imprecise word recognition impede higher-order processes, such as semantic retrieval and syntactic judgment, due to a lack of residual cognitive space. As a result, it leads to significantly reduced reading comprehension (Katz, Shankweiler, & Liberman, 1981; Stanovich, Siegel, & Gottardo, 1997). Because neighborhood candidates (e.g., phonology, spelling patterns, syntax, and semantic properties) are activated upon the stimulus of the text in the face of interference (Swanson, Zheng, & Jerman, 2009), effective activation and inhibition processes through the maintenance of pertinent memory traces are crucial while reading. Therefore, the efficiency of the informational filtering system through memory traces facilitates reading comprehension.

Due to its phonological processing and the maintenance of task-relevant information in an active state, verbal memory is robustly related to the performance of complex cognitive tasks, such as vocabulary acquisition and reading (Baddeley, 1986). Research has demonstrated that verbal short-term memory abilities and reading acquisition have a shared set of processes and utilize a common cognitive system (Adams & Gathercole, 1996; Gupta, 1996; Gupta & MacWhinney, Feldman, & Sacco, 2003; Henry & MacLean, 2003). A reliable relationship has been observed between digit span recall and word knowledge even when other variables, such as age and IQ scores, were controlled for (Gernsbacher, 1990; Gupta, 1996).

Gupta (1996) has attempted to explain the relation between immediate serial recall and word learning using a computational model in which a general sequencing mechanism provides immediate memory for the sequence of word forms. The model consists of three vital levels, including a phoneme layer, a phonological-chunk layer, and a semantic/context layer. The phoneme layer relates to output phonology at which phonemes are represented. The phonological-chunk layer entails the representations for word forms which are shared by input and output phonology. The last level, the semantic/context layer, represents semantic and contextual information about word forms (see Gupta, 1996, for details). According to Gupta's (1996) model which takes a new processing-oriented approach to examining word learning, the significant relationship between immediate serial recall and word learning lies in the common dependence of these two capabilities on core phonological and semantic processing mechanisms.

As a whole, verbal memory is indispensable with regard to reading in that textual inputs go through multiple processes, including encoding and retrieval of phonological, orthographic, and semantic referents stored in the mental lexicon (Gupta, 1996; Swanson, Saez, & Gerber, 2006). Through these processes, activated information about words, phrases or sentences should be sustained for a short period of time in order for the contents to be integrated into a context.

Daneman and Carpenter (1980) developed a reading-span test (RST) to investigate the contribution of working memory to reading comprehension. In the RST, participants read a few sentences aloud and were asked to remember the last word of each sentence. Due to the limited capacity of working memory, the mental resources to read a sentence and store the last word of each sentence were limited. Since participants needed to allocate mental resources efficiently during the task, the RST measured storing and processing capabilities concurrently upon reading. There was a significant correlation found between the RST score and reading comprehension performance (Daneman & Carpenter, 1980). A meta-analysis of studies using working memory-span tasks conducted by Daneman and Merikle (1996) also showed that working-memory-span capacity was correlated significantly with reading skills.

There are individual differences in working memory capacity and these differences are related, in part, to language comprehension (Daneman & Carpenter, 1983; Just & Carpenter, 1992). Daneman and Carpenter (1983) have suggested that the semantic processing of sentence comprehension is attributable to individual differences in memory capacity. Students who show efficient working memory retrieval (i.e., higher scores on the RST) are more likely to make use of fewer resources for the semantic processing of sentences, and, as a result, have sufficient resources to retain the words. In contrast, participants who show inefficient working memory (i.e., lower scores on the RST) have difficulty retaining target words due to the insufficient working memory capacity during reading. Similarly, from a capacity-oriented perspective, skilled readers make more text-based inferences because the multiple sources of processing, such as reading a sentence and storing the last word of each sentence at the same time, are available in working memory. Hence, high functioning students with greater working memory capacity are able to sustain more information necessary to complete the given task, such as reading, because they successfully utilize the efficient activation of semantic and syntactic information (Budd, Whitney, & Turley, 1995).

L1 and L2 Interdependence and the Korean Language

A myriad of studies have demonstrated cross-language links in the acquisition of language and reading skills, and been explained by the cross-linguistic interdependence hypothesis (Cummins, 1994). Phonological processing skills, for example, can transfer from one language to another (Gottardo, Yan, Siegel, & Wade-Woolley, 2001; Pae, Sevcik, & Morris, 2010). If one has robust L1[1] oral skills and reading proficiency, the likelihood for the child to gain success in L2 oral performance and reading is greater than those who lack those skills. Within the Roman alphabetic languages, children are able to make use of commonalities between L1 and L2, when they learn L2. Although universality exists in the acquisition of languages, language-specific variations are observed, because learning to read accommodates how a writing system maps phonological properties in the language under consideration (Pae, 2011; Perfetti & Liu, 2005). Therefore, it is possible that working memory and L1 ability may influence the learning of some second languages more than others, particularly if those languages use different forms of script or orthography.

The ability to control attention for the active maintenance of given information and the inhibition of irrelevant information is viewed as a domain-general construct (Kane et al., 2004; Payne, Kalibatseva, & Jungers, 2009). With robust evidence that verbal memory skills contribute to L1 reading comprehension (Daneman & Carpenter, 1980, 1983; Kane et al., 2004; Swanson & Berninger, 1995), the significant role of working memory has been expanded to the prediction of L2 language processing and reading comprehension. For instance, independent contributions of working memory for Hebrew-speaking high-school students were found in their learning English as L2 (Abu-Rabia, Share, & Mansour, 2003). A study of Japanese-speaking students evidenced a significant correlation in working memory between L1 and L2, as well as a mediator effect of L2 working memory on the relationship between L1 working memory and L2 syntactic comprehension (Miyake & Friedman, 1998). Significant relationships among variables of working memory, L1 reading comprehension, and L2 reading comprehension were also found in English-speaking college students' Spanish comprehension (Payne, Kalibatseva, & Jungers, 2009). A salient contribution of verbal working memory, measured using memory span and tongue twister, to text comprehension by Chinese children has also been reported (Leong, Tse, Loh, & Hau, 2008). The significant role of phonological memory has been expanded to L2 speech production for English-speaking adults who were learning Spanish as L2. O'Brien, Segalowitz, Collentine, and Freed (2006) found that phonological memory (serial nonword recognition) explained significant amounts of the variance in L2 narrative oral skills for both less proficient and more proficient adults, after controlling for speech output.

In cross-language research, a study of English and Korean offers an excellent opportunity to examine between-language interdependence because the two languages share the alphabetic principle (Pae, 2011; Ziegler & Goswami, 2005), but exhibit profound differences in their visual lexical form (linearity vs. block layout) as well as orthographic, phonological, and linguistic features. Korean is noticeably different from other languages in its origin; the

[1] Since L1 is to be processed automatically, the automaticity can be a way to determine the child's L1. In this study, a dominant language and L1 are operationalized as the language in which a child knows *instantly, intuitively, and effortlessly*, when he/she processes an utterance in a language, regardless of the first language to which the child was exposed. Hence, L1 was used throughout the paper for the sake of consistency with previous studies in order to refer to a dominant language.

Korean writing system was invented and promulgated by King Sejong in 1443, rather than being evolved over time. The key differences between English and Korean include lexicality and orthographic depth, writing systems, phonological differences, syntactic word order in sentences, and subject and/or object omissions in sentences.

English is considered to be a deep orthography (Katz & Frost, 1992), while Korean is a shallow orthography in which any words, even if unfamiliar words or nonwords, can be sounded out quickly and accurately. The Korean writing system, an alphabetic syllabary (Pae, 2011; Taylor, 1980, 1997), is unlike those of most other Roman alphabetic languages. The graphemes are not written one after another in a horizontal line, but form a square-like syllable block consisting essentially of a phonetic syllable (e.g., 한글 rather than ㅎㅏㄴㄱㅡㄹ). The number of Korean syllables currently in use in Korea is 2,457 (see Pae, 2011 for details), which is larger than Chinese which contains about 400 syllables or 1,300 syllables with tones (Taylor, 1997) and Japanese that includes fewer than 113 syllables (Taylor, 1997). However, the number of syllables is much smaller than that of English which includes about 8,000 syllables (Katz & Frost, 1992). Korean phonemes do not include the labio-dentals, interdentals, sibilants, or retroflexes that are found in English.

Korean falls into the Ural-Altaic family of language, with an agglutinative structure of grammar (i.e., nouns and verbs are formed by attaching derivational and/or inflectional prefixes and suffixes to the root). The basic syntactic structure of Korean, like Japanese, is a predicate-final structure with the basic word order of Subject-Object-Predicate (verb or adjective). In other words, Korean is a verb/predicate-final language (i.e., a language in which the verb or predicate always comes at the end of the sentence; Sohn, 1999). The Korean language is a context-rich language in that most of the sentences do not carry subjects, and, often times, objects are also omitted. The listener is forced to rely on the context in which the utterance takes place in order to comprehend the true meaning of the sentence. For example, "I love you" in English can be said "love" in Korean. With the omissions of the subject and the object, both the speaker and the listener are able to decipher the meaning of the utterance without problems in Korean by relying on the context in which communication occurs.

The Present Study

Despite the comparatively wide knowledge base on literacy acquisition in linearly arranged alphabetic languages, little is known about literacy development in writing systems with non-linear symbol arrangements. Korean is an outstanding example of alphasyllabic languages, as briefly noted earlier, because it entails the alphabetic principle (i.e., a grapheme maps onto a phoneme and strings of graphemes form a syllable) as well as a distinct visual syllabic structure with an adhesive rule of a consonant to a vowel. No consonant strings are allowed, except five doublets (ㄲ ㄸ ㅃ ㅆ ㅉ) and legal consonant-consonant forms at the bottom of the top-down syllable (see Pae, 2011, for more details).

Given that reading is a high-order cognitive function, the aim of this study was to investigate the independent contributions of three different types of verbal working memory to L2 reading fluency and comprehension via a comparison of English and Korean. Two research questions were posed as follows:

1. What is the individual contribution of phonological working memory, verbal digit working memory, and sentence verbal working memory to L1 and L2 reading for English-speaking bilinguals in the U.S. who learn Korean as L22?

[2] Again, for the sake of consistency, L1 and L2 were used in this study, although the U.S. children might have exposed to Korean first, regardless the dominance of language skills in English and Korean. The U.S. participants'

2. What is the individual contribution of phonological working memory, verbal digit working memory, and verbal sentence working memory to L1 and L2 reading for Korean-speaking bilinguals in Korea who learn English as L2?

On the basis of the findings of previous research, it was assumed that phonological working memory, measured using forward digit spans, has more of a working memory load than the sentence task for children, while verbal digit working memory has an extra transformational processing element and has the highest demand for working memory processes. Verbal sentence working memory includes a rich set of contextual and semantic information, allowing for long-term memory traces that should free up working memory to some extent. This task may have the least load on working memory process. Therefore, it was hypothesized that verbal digit working memory was more predictive for comprehension than phonological working memory and sentence recall skills.

Although the literature indicates a significant relationship between L1 verbal memory and L2 reading comprehension in Roman alphabetic languages, there has been little exploration of the role of verbal working memory in accounting for reading comprehension in an alphasyllabic language. We hypothesized that verbal memory capacity in L1 would be predictive of reading achievement in both L1 and L2, but the extent to which verbal memory played a role in English and Korean performance would be different when learning the language as L2. Because Korean has a consistent grapheme-phoneme correspondence, Korean involves direct lexical access in text processing with a little phonological mediation[3]. Given the grapheme-phoneme consistency in Korean, it was also hypothesized that the amount of variance explained by verbal memory would be smaller in Korean than that in English when learning the language as L2.

We also examined whether differences exist between English-speaking bilinguals and their Korean-speaking counterparts in terms of the strength of relationship between L1 and L2. This analysis would lead to a better understanding of L2 acquisition in learning contexts: one in an English-as-a-second-language (ESL)[4] context by English-Korean bilinguals and the other in an English-as-a-foreign-language (EFL) context by Korean-English bilinguals. If the language system had an influence on reading performance, verbal memory would play a role in a different way across sites where the L1 is different from each other. We expected a different pattern of relationship between verbal working memory and reading (i.e., different amount of the variance explained in the prediction model) across the two sites.

Methods

Participants

Fifty first- and second-grade children took part in this study. The participants consisted of two groups. The first group comprised 29 Korean immigrant children residing in the U.S. who

---

home language was predominantly Korean, although the children mostly communicate in English at school and at home in which Korean was spoken by the participant's parent.

[3] The dual-route hypothesis, including the lexical route (a.k.a., addressed, word-detector, or lexical look-up route) and the sublexical route (a.k.a., nonlexical, assembly, rule-based, or phonologically mediated procedure), offers an explanation of mapping print into sound. Since the word processing route was beyond the scope of this study, the two routes were not considered in this study.

[4] Although it can be questionable as to whether the Korean-immigrant children in the U.S. were actually learning Korean as an L2-equivalent language, the questionnaire by participant's parent and a brief interview with the child pointed to language automaticity established in English. See Results for further information on the language dominance.

were learning Korean as L2 (English-Korean bilinguals) within a Korean educational program (i.e., a Saturday school) in a metropolitan area in the southeastern United States. The second group was composed of 21 Korean natives who lived in Korea and learned English as L2 (Korean-English bilinguals) in the formal education setting in Busan, South Korea. In order to match the socioeconomic status across the sites, the participants in South Korea were restricted to students who were attending a private elementary school. Parental informed consent and child assent were obtained for all the participants.

The English-Korean learners' mean age was 91.82 months (SD = 8.62; 15 males, 14 females). The Korean-English counterparts' mean age was 87.66 months (SD = 5.92; 13 males, 8 females).

The parent questionnaire and the child oral report indicated that all the participants were not exposed to other languages beyond the two languages under consideration. It should be noted that the English-Korean bilingual children in the U.S. might have been exposed to Korean first because their parents spoke Korean at home. The questionnaires completed by the parent and brief interviews with the children indicated that the parent typically spoke to the child in Korean, and the child responded to his/her parent in English. Hence, the strength of their language skills was shown in English, and their oral production in Korean was laborious, which was a typical indication of L2. Because of the wide range of variability in both home-language use and language dominance for the English-Korean speaking bilinguals, the U.S. participants' dominant language was validated using object picture naming latency in L1 and L2. This validation procedure was derived from the premise that object naming in L1 is typically faster than L2, due to the additional processing demand to resolve a competition from L1 candidates (Kroll & Stewart, 1994). The results showed that the participants took significantly longer naming objects in L2 than L1 [U.S. participants: $t(1,28) = 7.70$, $p < .00$], validating the participants' and parents' self-reports. Since the Korean native students spoke Korean as their L1, the validation procedure was not necessary.

Procedure

A test battery, which included measures of forward and backward digit recall, sentence recall, and reading skills, was individually administered at the Korean schools or the participant's home by a bilingual examiner. Before the test administration, a brief interview with the child was performed, in addition to child assent.

Measures

Based on the protocols of English measures which were U.S. norm-referenced, the Korean measures were experimentally designed to achieve comparability with the English version. Word frequency, face validity, inter-item consistency, and test-retest reliability were taken into account to overcome potential limitations and to maximize the cross-cultural comparability. For the experimental Korean measures, internal consistency and test-retest reliability coefficients ranged from .79 to .91.

Predictor Measures

Phonological Working Memory. The Forward Digit Span subtest of the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) was used as a measure of phonological working memory. The forward digit span task required an immediate recall of auditory number strings forwards at a rate of one digit per second and asked the child to simply repeat the list of digits as heard. This task involves a phonological memory reservoir that stores in-coming information passively for a few seconds, and then produces articulatory output under constraints.

Verbal Digit Working Memory. The Backward Digit Span subtest of the WISC-III (Wechsler, 1991) was used as a measure of verbal digit memory. This test asked the child to say the list of numbers in a reversed order for an instant recall. Since it was designed to measure the participant's ability to store and manipulate digits that were in temporary short-term memory, this task entailed an extra-transformational processing element. It required the child not only to recite the presented information, but also to make some transformation. The additional processing on top of memory makes this task quite powerful and consistent predictor of a wide array of high order processes related to fluid intelligence, such as reading comprehension.

For the forward and backward digit span tests, a set of two lists with the same length was provided, and if at least one of the two lists was correctly produced, the subsequent list was increased by one digit until the child reached the ceiling. The internal consistency values and test-retest reliability coefficients for age groups of 6 to 9 were high, ranging from .89 to .94.

The same protocol was used for the Korean measure, but the digits were presented in Korean by the examiner and the participants responded in Korean as well. Since the articulation of the number in a language involves its phonological coding, using the same stimuli in the two languages maintains a unitary construct.

Verbal Sentence Short-Term Memory. The Sentence Repetition subtest of *A Developmental Neuropsychological Assessment* (Korkman, Kirk, & Kemp, 1998) assessed the ability to recall sentences of increasing complexity and length. The examinee was asked to repeat verbatim the sentence presented by the examiner. The generalizability coefficients for children of 6 to 9 years of age ranged from .72 to .85. The stability coefficient was .91 for that age group. The Korean measure of sentence repetition was constructed on the basis of the English version, by attending to the word and sentence lengths and the semantic complexities of the given sentences. The internal consistency coefficient was .81 for the Korean sentence repetition test.

Outcome Measures

Reading Fluency. The Fluency subtest of the Gray Oral Reading Test-4 (GORT-4; Wiederholt & Bryant, 2001) was employed as a reading fluency assessment tool. The GORT-4 is composed of 4 subtests, including reading rate, accuracy, fluency, and comprehension; the fluency score is a composite score of rate and accuracy scores. The GORT-4 was designed to determine the particular reading strengths and weaknesses that individual students possess and to serve as a measurement device in assessing reading ability (Wiederholt & Bryant, 2001).

Reading Comprehension. The Comprehension subtest of the Gray Oral Reading Test-4[5] (GORT-4; Wiederholt & Bryant, 2001) was employed as a reading comprehension assessment tool. The GORT-4 evidences a high degree of reliability, which was consistently high across all three types of reliability: content sampling ranged from .91 to .97; test-retest, .85 - .95; and scorer differences, .94 - .99. For the Korean version of the reading fluency and comprehension measures, Form B stories of the GORT-4 were translated into Korean so that the participants could receive an equivalent fluency and comprehension measures in the two languages.

---

[5] There have been concerns about content validity and concurrent validity of the GORT-4 (Keenan & Betjemann, 2006). However, Keenan & Betjemann (2006) administered the GORT passages to college students to examine the extent to which GORT questions were passage-independent items. Since our participants' ages fell in the range of the GORT-4' normative sample which was 6-18 years of age, we followed the GORT manual.

Results

Descriptive Statistics and Correlations

A descriptive analysis of all data was initially conducted, including an examination of variable means and standard deviations, outlier checks, and the distribution of scores. Since the focal point centered on reading fluency and comprehension in relation to verbal memory, word and nonword identification skills were not included as variables in this study. The distribution of the raw scores on the reading tests was slightly positively and negatively skewed according to the measures, but no outstanding differences were found for the planned inferential statistical analyses. Due to cross-cultural measurement issues given that no measures utilized were normed in Korea, raw scores were used for analyses. As expected, the participants consistently demonstrated higher performance in their L1 than L2, which was a typical pattern found in previous research (Pae, Sevcik, & Morris, 2004, 2010). The two groups did not differ in age as assessed by a *t*-test corrected for unequal variances, $t = 1.91$, $p > .05$. Although the U.S. children slightly outperformed the Korean counterparts, the children's performance on the measures was not significantly different for the U.S. and Korean groups, except for L1 sentence repetition and L1 backward digit span ($F = 8.17$, $p < .01$; $F = 9.90$, $p < .01$, respectively). Table 1 shows the means and standard deviations for L1 and L2 measures. The reading measures, including fluency and comprehension, did not differ between the two groups.

Table1. *Descriptive Statistics*

	English-Korean Bilinguals			*Korean-English Bilinguals*		
	Mean	*SD*	*Range*	*Mean*	*SD*	*Range*
L1 Forward Digit Span	7.69	1.83	5-11	7.19	1.57	5-10
L2 Forward Digit Span	5.90	1.49	4-9	6.05	1.32	4-8
L1 Backward Digit Span	5.55	1.84	2-9	4.67	1.15	3-7
L2 Backward Digit Span	3.83	1.23	2-6	3.71	1.27	2-6
L1 Sentence Repetition	16.72	3.99	9-26	13.57	3.64	8-20
L2 Sentence Repetition	10.10	4.04	3-19	8.81	2.44	6-14
L1 Reading Fluency	40.69	17.39	1-71	40	15.93	16-84
L2 Reading Fluency	8.10	7.52	0-25	11.90	7.65	0-35
L1 Reading Comprehension	20.28	11.09	0-42	16.10	8.65	5-40
L2 Reading Comprehension	6.21	3.89	0-13	6.81	4.02	0-18

Since each participant was tested on L1 and L2 skills across the two sites, and provided scores for each permutation of the variables, ANOVA was performed with two foci. The first focus was placed on the difference between the learner groups (i.e., ESL U.S. vs. EFL Korea) to examine main effects and interactions with site (U.S. vs. Korea) as the between-subject factor and language (L1 vs. L2) as the within-subject factor. The results revealed a main effect of language ($F(1, 48) = 42.21$, $p = .000$, partial $\eta^2 = .32$), indicating that there are differences in the means of reading fluency skills between the two learner groups when ignoring other factors. There was an interaction effect, indicating that the children's performance was not the same at the two sites [$F(1,48) = 5.27$, $p = .026$, partial $\eta^2 = .10$]. Main effects for sentence repetition, forward digit span, and backward digit span were significant. There were interaction effects for forward and backward digit span [$F(1, 48) = 4.45$, $p = .04$, partial $\eta^2 = .09$; $F(1, 48) = 4.38$, $p = .04$, partial $\eta^2 = .08$, respectively], suggesting that the verbal memory skills in digit span were not uniform. The two-way interaction between the two sites for the sentence repetition task did not reach significance [$F(1, 48) = 3.52$, $p = .07$, partial $\eta^2 = .07$].

For the reading measures, there were significant main effects for reading fluency and reading comprehension performance. A markedly significant interaction implied that this pattern of variance was different for the two participant groups.

The second focus was placed on the language command (i.e., L1 vs. L2). Since the language command was a within-subject variable, the main effect and interaction effect were not reported. The L1 backward digit span and L1 sentence repetition were significantly different across language commands ($F(1,48) = 9.90$, $p = .003$, $F(1,48) = 8.17$, $p = .006$, respectively).

Zero-order correlation coefficients among the variables were obtained. Table 2 reports the bivariate correlation coefficients in the lower-left triangle and the partial correlation coefficients controlling for age in the upper-right triangle for the English-speaking participants. The partial correlation coefficients were computed because verbal memory is age-sensitive. For the U.S. students, there were moderate to high significant correlations among the variables under consideration after controlling for age (r ranges: .41 - .85, $p < .05$). Interestingly, the sentence repetition performance showed the lowest correlations with the other variables.

For the Korean counterparts, the pattern of the significant correlations among the variables was different from that of the English-speaking participants (see Table 3). Korean (L1) sentence recall proficiency was significantly correlated with Korean (L1) and English (L2) backward digit span (r = .52 and .57, respectively) after controlling for age. Overall, the Korean natives' scores showed comparatively lower correlations among the variables under consideration than those of the U.S. children.

Table 2. *Correlations among the Variables below the Diagonal and Partial Correlations Controlling for Age above the Diagonal for the English-Speaking Children in the U.S.*

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1. L1 Forward Digit Span	1	.85***	.51**	.43*	.62***	.38*	.65***	.47*	.28	.49**
2. L2 Forward Digit Span	.84***	1	.45*	.34	.48*	.49**	.59**	.43*	.23	.56**
3. L1 Backward Digit Span	.51**	.46*	1	.59**	.50**	.32	.64***	.56**	.62***	.54**
4. L2 Backward Digit Span	.44*	.36	.61**	1	.44*	.23	.45*	.41*	.51**	.44*
5. L1 Sentence Repetition	.61***	.47*	.45*	.41*	1	.54	.63***	.36	.36	.15
6. L2 Sentence Repetition	.35	.45*	.23	.17	.54**	1	.49**	.43*	.37	.36
7. L1 Reading Fluency	.65***	.59**	.64***	.46*	.62**	.44*	1	.71***	.70***	.64***
8. L2 Reading Fluency	.47*	.42*	.52**	.40*	.37	.42*	.70***	1	.70***	.79***
9. L1 Reading Comprehension	.25	.19	.48**	.42*	.36	.41*	.63***	.68***	1	.64***
10. L2 Reading Comprehension	.49**	.57**	.55**	.45*	.14	.32	.64***	.78***	.58**	1

*$p < .05$; ** $p < .01$; *** $p < .001$

Table 3. *Correlations among the Variables below the Diagonal and Partial Correlations Controlling for Age above the Diagonal for the Korean-Speaking Children in Korea*

	1	2	3	4	5	6	7	8	9	10
1. L1 Forward Digit Span	1	.48*	.31	.54*	.37	.58**	-.03	.58**	-.06	.50*
2. L2 Forward Digit Span	.57**	1	.48*	.70**	.40	.36	-.02	.30	.18	.36
3. L1 Backward Digit Span	.39	.57**	1	.63**	.52*	.66**	-.13	.23	-.27	.07
4. L2 Backward Digit Span	.61**	.75***	.68**	1	.57**	.50*	.16	.44	.12	.42
5. L1 Sentence Repetition	.47*	.56**	.60*	.64**	1	.54*	.06	.22	-.15	.09
6. L2 Sentence Repetition	.64**	.50*	.70***	.58**	.63**	1	-.04	.59**	-.17	.39
7. L1 Reading Fluency	.08	.15	-.02	.25	.18	.08	1	.16	.84***	.29
8. L2 Reading Fluency	.64**	.45*	.33	.52*	.36	.65**	.25	1	.03	.87***
9. L1 Reading Comprehension	-.03	.19	-.23	.14	-.10	-.13	.83***	.05	1	.29
10. L2 Reading Comprehension	.51*	.36	.10	.43	.13	.40	.31	.84***	.30	1

*$p < .05$; ** $p < .01$; *** $p < .001$

Hierarchical Regression Analyses for Reading Fluency and Comprehension

In order to explore the respective contribution of the phonological and verbal memory in L1 and L2 reading skills, a series of hierarchically nested regressions were performed separately for each learner group. After controlling for age, forward digit span capacity was entered first, backward digit span second, and then sentence repetition was entered last into a model as predictor variables. Reading fluency and comprehension skills served as dependent variables. The within-language analyses of L1-L1 and L2-L2 as well as the between-language analysis of L1-L2 were performed. The results of hierarchical regression analyses for the English-speaking children are shown in Table 4.

The English forward digit span accounted for 41% of the unique variance in English reading fluency (F change (1,26) = 19.14, p = .000), and the English backward digit span explained an additional 13% of the variance in English reading fluency (F change (1,25) = 7.04, p = .014). The English backward digit span was predictive of 29% of the unique variance of English reading comprehension (F change (1,25) = 12.53, p = .002). The L2 Korean forward digit span was also a significant predictor of L2 Korean fluency and comprehension skills (17% and 31% of the unique variance, respectively). When it came to cross-language prediction, the English forward and backward digit span tasks explained the unique variances in both Korean reading fluency and reading comprehension. The English (L1) forward digit span task explained 23% and the English backward digit span measure accounted for 13% of the additional unique variance in Korean (L2) reading fluency (F change (1,26) = 7.54, p = .010; F change (1,25) = 5.35, p = .029, respectively). The English (L1) forward span task explained 23% of the unique variance and the backward digit span test accounted for an additional 11% of the unique variance in Korean (L2) reading comprehension (F change (1,26) = 8.26, p = .008; F change (1,25) = 4.31, p = .048, respectively).

Table 4. *Hierarchical Regression Models for English-Korean Bilinguals in the U.S.*

Dependent Variables	Predictors	R^2	ΔR^2	β	t
Within Language (English-English)					
L1 Fluency	Step 1. Age	.02	.02	ns	ns
	Step 2. L1 Forward Digits	.43***	.41***	.44	2.81*
	Step 3. L1 Backward Digits	.56***	.13*	.44	2.65*
	Step 4. L1 Sentences	.60***	.04	ns	ns
L1 Comprehension	Step 1. Age	.07	.07	.46	2.73*
	Step 2. L1 Forward Digits	.14	.07	ns	ns
	Step 3. L1 Backward Digits	.43**	.29**	.63	3.21**
	Step 4. L1 Sentences	.44**	.01	ns	ns
Within Language (Korean-Korean)					
L2 Fluency	Step 1. Age	.00	.00	ns	ns
	Step 2. L2 Forward Digits	.18	.18*	ns	ns
	Step 3. L2 Backward Digits	.26	.08	ns	ns
	Step 4. L2 Sentences	.32	.06	ns	ns
L2 Comprehension	Step 1. Age	.02	.02	ns	ns
	Step 2. L2 Forward Digits	.33**	.31**	.42	2.22*
	Step 3. L2 Backward Digits	.39**	.06	ns	ns
	Step 4. L2 Sentences	.40*	.01	ns	ns
Between Languages (English-Korean)					
L2 Fluency	Step 1. Age	.00	.00	ns	ns
	Step 2. L1 Forward Digits	.23*	.23*	.48	7.54*
	Step 3. L1 Backward Digits	.36*	.13*	.46	2.20*
	Step 4. L1 Sentences	.36*	.00	ns	ns
L2 Comprehension	Step 1. Age	.02	.02	ns	ns
	Step 2. L1 Forward Digits	.25*	.23**	.48	2.36*
	Step 3. L1 Backward Digits	.36**	.11*	.51	2.64*
	Step 4. L1 Sentences	.45**	.09	ns	ns

Note * $p < .05$; ** $p < .01$; *** $p < .001$
ns = not significant

As seen in the correlation matrix, a difference of the prediction was also found in the regression analysis for the Korean-speaking children from the English-speaking counterparts (see Table 5). As hypothesized, the predictive power of the phonological and verbal memory skills in reading fluency and comprehension diminished for the Korean-speaking children. In the Korean (L1) within-language relationship, no predictors accounted for the significant variance in Korean reading fluency and comprehension. The English (L2) sentence repetition task explained 15% of the unique variance in English reading fluency (F change (1,16) = 4.59, p = .048). The Korean (L1) forward digit span explained a unique variance of 28% in English (L2) reading fluency and an additional 24% of the variance in L2 reading comprehension.

Table 5. *Hierarchical Regression Models for Korean-English Bilinguals in Korea*

Dependent Variables	Predictors	R^2	ΔR^2	β	t
Within Language (Korean-Korean)					
L1 Fluency	Step 1. Age	.08	.08	ns	ns
	Step 2. L1 Forward Digits	.08	.00	ns	ns
	Step 3. L1 Backward Digits	.09	.01	ns	ns
	Step 4. L1 Sentences	.11	.02	ns	ns
L1 Comprehension	Step 1. Age	.01	.01	ns	ns
	Step 2. L1 Forward Digits	.01	.00	ns	ns
	Step 3. L1 Backward Digits	.08	.07	ns	ns
	Step 4. L1 Sentences	.08	.00	ns	Ns

Hierarchical Regression Models for Korean-English Bilinguals in Korea

Within Language (English-English)					
L2 Fluency	Step 1. Age	.15	.15		
	Step 2. L2 Forward Digits	.23	.08		
	Step 3. L2 Backward Digits	.32	.09		
	Step 4. L2 Sentences	.47*	.15*	.49	2.14*
L2 Comprehension	Step 1. Age	.02	.02		
	Step 2. L2 Forward Digits	.14	.12		
	Step 3. L2 Backward Digits	.19	.05		
	Step 4. L2 Sentences	.24	.04		
Between Languages (Korean-English)					
L2 Fluency	Step 1. Age	.15	.15		
	Step 2. L1 Forward Digits	.43**	.28**	.56	2.55*
	Step 3. L1 Backward Digits	.44*	.01		
	Step 4. L1 Sentences	.44*	.00		
L2 Comprehension	Step 1. Age	.02	.02		
	Step 2. L1 Forward Digits	.26	.24*	.58	2.34*
	Step 3. L1 Backward Digits	.27	.01		
	Step 4. L1 Sentences	.28	.01		

Discussion

The aim of this study was to investigate the extent to which phonological and verbal memory capacity, as indicated by forward and backward digit spans and sentence immediate recall tasks, affected reading outcomes in English and Korean by English-Korean and Korean-English bilinguals. The general pattern of L1 and L2 performance by the participants in the two international sites was consistent with the findings of previous research, indicating that the students demonstrated a greater strength in L1 than L2 (Pae, Sevcik, & Morris, 2004, 2010). The pattern of the bilinguals' performance on the given measures was different across the two sites. The difference in reading outcomes might result from the L2 learning context; one with an ESL context and the other with an EFL context.

Obviously, the EFL students in Korea, in which the Korean language is ubiquitously spoken, have limited L2 exposure, input, and use, compared to their ESL counterparts in the U.S. whose dual languages are spoken on a daily basis. The English-speaking children in the U.S. appeared to perform better in phonological and verbal memory tasks. Since phonological and verbal memory is closely linked to the phonological representations of the language, the richer phonological component of English (due to the number of legal sequences in speech sounds) than Korean might have influenced the English-speaking children's verbal memory performance. As stated earlier, the number of syllables in English and Korean is drastically different (2,000 vs. 8,000). This finding suggests that the availability of a wide range of phonological information may facilitate the maintenance and manipulation of input information and eventually help the production of the stored phonological information. However, further research is warranted for the explanation of the role of phonological and verbal memory in reading across different languages.

The partial correlations, controlling for age, showed significant correlations between L1 and L2. This result is consistent with the notion of cross-language interdependence (Cummins, 1994). There were significant correlations between phonological memory and verbal working memory and between these skills in L1 and L2, suggesting that the two indicators reflect similar constructs. Interestingly, the sentence verbal working memory demonstrated low correlations with other variables for the English-speaking children in the U.S. These low correlations indicate that the sentence repetition task measured the same qualities as other variables to some extent, but the overlapping is not conspicuous, suggesting that the sentence repetition task might measure a unique element of verbal memory skills.

No significant correlations were found for the Korean-speaking participants residing in Korea in between-language reading fluency and comprehension. This can be explained as the comparatively limited proficiency of L2 fluency and comprehension skills, compared to the acquired optimal level of L1 skills. The magnitude of the correlation coefficients among the variables was larger in the English L1 group in the U.S. One possible explanation for the weaker correlations in the Korean L1 children may be the fact that effects stemming from digit and sentence recall can mask the reading outcomes of L1 and L2 due to fewer demands imposed on phonological coding resulting from the shallow orthographic nature of Korean (Pae, 2011). As Gupta (1996) notes, phonological mapping appears to be related to the interaction between verbal memory and reading-related activities.

The Individual Contribution of Phonological and Verbal Memory to Reading Fluency and Comprehension for the English-Speaking Children (Research Question 1)

For the English-speaking children who were learning Korean as L2 in the U.S., phonological memory capacity, measured using the forward digit span task, played a salient role in English and Korean within-language reading fluency and comprehension. It also played a significant role in the prediction of cross-language reading fluency and comprehension. Digital verbal working memory, as indicated by the backward digit span task, played a significant role in English within-language reading fluency and comprehension as well as cross-language reading fluency and comprehension.

The pattern of findings can be explained in four ways. First, the weak relationship between verbal memory and reading comprehension in Korean can be attributable to the shallow orthography in which the grapheme-phoneme correspondence is transparent and regular. Since learners can easily decode words without problems, reading Korean may not necessitate a significant memory load for fluent reading. Second, reading English may be more challenging because English has a wider phonological repertoire than Korean.

Specifically, the orthographically deep characteristic affects English reading at the phonemic level, as a single grapheme can be pronounced differently according to neighboring graphemes. Third, English words have richer semantic representations than Korean, and are multidimensional in that many words carry multiple meanings. For example, English words have many homographs (e.g., *bat*, a baseball vs. *bat*, an animal) and homophones (e.g., *hi* vs. *high*; *whole* vs. *hole*, and heteronyms (e.g., *The wind is too strong* vs. *Wind the toy, and it will move*) (Pae, Greenberg, & Williams, 2011). Hence, the ability to activate task-relevant information and to inhibit incongruent information with the given syntactic and semantic coherence may be directly related to the ability to recall orally presented stimuli. This capability may also aid the child to build extended meanings on the basis of the root word and strengthen his/her word repertoire, because word knowledge can be broadened through a facilitation of effective verbal memory capability and a core phonological and semantic processing mechanism (Gupta, 1996). Lastly, the relationship between cross-language verbal working memory and reading seems to be more challenging than that in L1 and L2 within languages. Specifically, both L1 phonological and verbal working memory played a significant role in L2 reading fluency and comprehension for the English-speaking children. Since L2 has weaker connections to the mental lexicon and storage (Kroll & Stewart, 1994), higher cognitive loads and demands may be required for L2 than L1.

The Individual Contribution of Phonological and Verbal Memory to Reading Fluency and Comprehension for the Korean-Speaking Children (Research Question 2)

The results of hierarchical regression analyses showed different predictive patterns across sites. In general, the capability of recalling the immediate serial numbers dominantly served as a significant predictor for the English-speaking children. However, the magnitude of association between verbal memory capacity and reading fluency and comprehension diminished for the Korean-speaking counterparts. In Korean within-language reading, the predictor variables did not account for a significant variance in reading fluency and comprehension. For English within-language reading, phonological working memory explained only English reading fluency. The ability to form phonological representations of serially presented number stimuli seems to be important to read fluently in English. This may be because skills in phonological coding and quick retrieval facilitate word reading (Gathercole & Baddeley, 1989).

In a similar vein to the finding with the English-speaking children, cross-language reading seems to be more challenging for the Korean-speaking children. The L1 phonological working memory skills played a significant part in both L2 reading fluency and comprehension. This result indicates that reading fluency and comprehension development in L2 English is contingent upon children's ability to recall and efficiently retrieve phonological input. Since the representations of words stored in an individuals' memory are multifaceted, the ability to make use of the connections between phonological information and the given sentence may vary across learners.

Reading English and Korean

Variability in the ability to retain and manipulate serially presented digits seems to reflect the likelihood of success in reading fluency and comprehension because reading requires interrelated process skills, such as accessing stored information, selecting a relevant meaning on the basis of contextual information, and evaluating the appropriateness of the chosen meaning. Since the pronounceability of a new word is vital in word learning, phonological working memory seems to play a dominant role in L2 reading skills for Korean native children.

One possible explanation about the different dimension across the learner groups may be the linguistic features of the two languages that the learner groups speak on a daily basis. It is possible that the demand of phonological coding depends on the extent to which English requires in the process of grapheme-phoneme mapping. Attempts have been made in the literature to examine the features of lexical structures influencing visual processing or phonological mediation (Coltheart, Laxon, Keating, & Pool, 1986; Rastle & Coltheart, 1999). A shallow orthography, such as Korean, involves automatic and direct access to stored orthographic and phonological representations, while a deep orthography, such as English, goes through a phonologically mediated procedure (Stone & Van Orden, 1993). As such, the results of this study can be aligned with the explanation of direct access or phonologically mediated access to the mental storage in that reading English involves stronger phonological mediation due to the nature of a deep orthography than reading Korean (see Coltheart, Laxon, Keating, & Pool, 1986; Rastle & Coltheart, 1999 for details). Besides, verbal memory permits learners to store input sounds in a manner that allows for easy storage and retrieval from long-term memory depending on the linguistic characteristics in which children learn.

To summarize, developing L2 fluency and comprehension is a long-term process and a multifaceted construct, which involves phonological coding, encoding, appropriate verbal working memory processing, and semantic decoding. Due to the interplay of the variables, the function of verbal working memory and reading fluency and comprehension cannot be disregarded in L1 and L2 reading.

Limitations and Future Directions

The present study was unique with respect to a contrast between a Roman alphabetic language and the alphasyllabic language as well as an employment of two learning contexts (i.e., ESL and EFL contexts), while investigating the role of verbal memory in reading fluency and comprehension in L1 and L2. However, the following limitations are worthwhile to mention. First, the small sample size limited the statistical power of the inferential analyses. Second, the lack of standardized instruments across sites deter the generalization of this study's findings. It is necessary to have culturally and linguistically appropriate instruments in cross-language studies. Third, cultural variations in the two sites were not taken into account in the analyses and interpretation. Fourth, the potentially confounding effects of phonological complexity in English and Korean on the task were not examined. As the phonological and phonotactic structure of words is language-specific, the articulatory difficulty at the boundaries of the number of words pronounced in sequence needs to be taken into account. The phonological complexity of the numerals might have influenced the children’s performance on the forward and backward digit span tasks. Lastly, there are still unanswered questions regarding the locus, nature, causality, and function of verbal memory in the development of L1 and L2 fluency and comprehension. A more systematic large-scale study would allow for an investigation to answer remaining questions.

Future studies examining the following are recommended. First, studies with a larger sample size will provide higher statistical power that will validate the findings of this study. Second, cross-cultural measurement development is needed to address complexities of two cultures, including cultural, contextual, linguistic differences, and to have valid psychometric properties to achieve measurement equivalence across the two groups. Third, since this study is a cross-sectional study, it cannot explain a developmental trajectory of the bilingual children in the two sites. A longitudinal study would offer an examination of children's dual-language learning and developmental trajectories in a systematic way.

• • •

Acknowledgement
Partial support for this study was provided to Hye Pae by the University Research Council of the University of Cincinnati. We thank the families of our participants for their cooperation and support. We also thank the anonymous reviewers for the valuable comments on the earlier version of this paper.

Hye K. Pae, Ph.D., is Assistant Professor of Literacy and Second Language Studies at the University of Cincinnati. Her research interests include psycholinguistics, second language acquisition, and assessment challenges across cultures.

Rose A. Sevcik, Ph.D., is Professor of Psychology at Georgia State University, Atlanta. Her research has focused on the language and reading acquisition of children and youth at risk for or with learning disabilities.

References

Adams, A. M., & Gathercole, S. E. (1996). Phonological working memory and spoken language development in young children. *The Quarterly Journal of Experimental Psychology, 49A*, 216-233.

Abu-Rabia, S., Share, D., & Mansour, M. S. (2003). Word recognition and basic cognitive processes among reading-disabled and normal readers in Arabic. *Reading and Writing: An Interdisciplinary Journal, 16*, 423-442.

Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.

Baddley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.). *The psychology of learning and motivation* (Vo. 8, pp. 47-89). New York: Academic.

Budd, D., Whitney, P., & Turley, K. J. (1995). Individual differences in working memory strategies for reading expository text. *Memory & Cognition, 23*, 735-748.

Coltheart, V., Laxon, V. J., Keating, G. C., & Pool, M. M. (1986). Direct access and phonological encoding processes in children's reading: Effects of word characteristics. *British Journal of Educational Psychology, 56*, 255-270.

Colom, R., Flores-Mendoza, C., Quiroga, M. A., & Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality & Individual Differences, 39,* 1005-1014.

Cummins, J. (1994). The acquisition of English as a second language. In R. Pritchard & K. Spangenberg-Urbschat (Eds.) *Kids come in all languages*. (pp. 36-62). Newark, DE: International Reading Association.

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

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9,* 561-583.

Daneman, M. & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3,* 422-433.

Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry, 39,* 3-27.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28, 200-213.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hillsdale, NJ: Lawrence Erlbaum.

Gernsbacher, M. A. (1990). Less skilled readers have less efficient suppression mechanisms. *Psychological Science, 4,* 294-298.

Gottardo, A., Yan, B., Siegel, L. S., & Wade-Woolley, L. (2001). Factors related to English reading performance in children with Chinese as a first language: More evidence of cross-language transfer of phonological processing. *Journal of Educational Psychology, 93 (3),* 530-542.

Gupta, P. (1996). Word learning and verbal short-term memory: A computational account. *Proceedings of the 18th Annual Conference of the Cognitive Science Society,* 189-194.

Gupta, P., MacWhinney, B., Feldman, H. M., and Sacco, K. (2003). Phonological memory and vocabulary learning in children with focal lesions. *Brain and Language, 87*, 241-252.

Henry, L. & MacLean, M. (2003). Relationships between working memory, expressive vocabulary and arithmetical reasoning in children with and without intellectual disabilities. *Educational and Child Psychology,* 20(3*), 51-63.*

Jeffries, S., & Everatt, J. (2004). Working memory: Its role in dyslexia and other specific learning difficulties. *Dyslexia: An International Journal of Research and Practice, 10*, 196-214.

Johnston, R. S., Rugg, M. D., & Scott, T. (1987). Phonological similarity effects, memory span and developmental reading disorders: The nature of the relationship. *The British Journal of Psychology, 78,* 205-211.

Jorm, A. F., Share, D. L., Maclean, R., & Matthews, R. (1984). Phonological confusability in short-term memory for sentences as a predictor of reading ability. *The British Journal of Psychology, 75,* 393-400.

Just, M. A., & Carpernter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122-49.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189-217.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67-84). Amsterdam: North Holland.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology, 32*, 474-484.

Keenan, J. M., & Betjemann, R. S. (2006). Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading, 10 (4),* 363-380.

Korkman, M., Kirk, U., & Kemp, S. (1998). *A Developmental Neuropsychological Assessment.* San Antonio, TX: The Psychological Corporation.

Kroll, J., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory & Language, 33,* 149-174.

Leong, C. K., Tse, S. K., Loh, K. Y., & Hau, K. T. (2008). Text comprehension in Chinese children: Relative contribution of verbal working memory, pseudoword reading, rapid automatized naming, and onset-rime phonological segmentation. *Journal of Educational Psychology, 100*, 135-149.

McDougall, S., Hulme, C., Ellis, A., & Monk, A. (1994). The role of short-term memory and phonological skills in reading. *Journal of Experimental Child Psychology, 58*, 112-133.

Miller, P., & Kupfermann, A. (2009). The role of visual and phonological representations in the processing of written words by readers with diagnosed dyslexia: Evidence from a working memory task. *Annals of Dyslexia, 59,* 12-33.

Miyake, A., & Friedman, N.P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A.F. Healy & L.E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Brien, I., Segalowitz, N., Collentine, J., Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics, 27*, 377-402.

Pae, H. K. (2011). Is Korean a syllabic alphabet or an alphabetic syllabary? *Writing Systems Research,* DOI: 10.1093/wsr/wsr002.

Pae, H. K., Sevcik, R., & Morris, R. D. (2004) Cross-language links between English and Korean in the second-language reading acquisition, *Journal of Multilingual Communication Disorders, 2 (1),* 68-79.

Pae, H. K., Sevcik, R. A., & Morris, R. D. (2010). Cross-language correlates in phonological process and naming speed: Evidence from deep and shallow orthographies. *The Journal of Research in Reading, 33(4),* 335-436.

Pae, H. K., Greenberg, D., & Williams, R. S. (2011). An analysis of differential error patterns on the Peabody Picture Vocabulary Test-III in children and struggling African-American adults readers. *Reading and Writing: An Interdisciplinary Journal*. DOI: 10.1007/s11145-011-9315-x.

Payne, T. W., Kalibatseva, Z., & Jungers, M. K. (2009). Does domain experience compensate for working memory capacity in second language reading comprehension? *Learning and Individual Differences, 19,* 119-123.

Perfetti, C. A. and Liu, Y. (2005). Orthography to phonology and meaning: comparisons across and within writing systems. *Reading and Writing: An Interdisciplinary Journal, 18*, 193-210.

Rastle, K. and Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception & Performance*, 25, 482-503.

Sohn, H-M. (1999). *The Korean Language*. New York: Cambridge University Press.

Stanovich, K. E., Siegel, L. S., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of Educational Psychology, 89 (1),* 114-127.

Stone, G., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 744-774.

Swanson, H. L., & Berninger, V. (1995). The role of working memory in skilled and less skilled readers' comprehension. *Intelligence, 21,* 83-108.

Swanson, H. L., Saez, L., & Gerber, M. (2006). Growth in literacy and cognition in bilingual children at risk or not at risk for reading disabilities. *Journal of Educational Psychology, 98*, 247-264.

Swanson, H. L., Zheng, X., & Jerman, O. (2009). Working memory, short-term memory, and reading disabilities: A selective meta-analysis of the literature. *Journal of Learning Disabilities, 43 (3),* 260-287.

Taylor, I. (1980). The Korean writing system: An alphabet? a syllabary? a logography? In P. A. Kolers, M. E. Wrostad & H. Bouma (Eds.). *Processing of visible language 2.* New York: Plenum Press.

Taylor, I. (1997). Psycholinguistic reasons for keeping Chinese characters in Korean and Japanese. In H. C. Chen (Ed.), *Cognitive processing of Chinese and related Asian Languages* (pp. 299-319). Hong Kong: Chinese University Press.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*, 104-132.

Wechsler, D. (1991). *The Wechsler intelligence scale for children-III*. New York: Psychological Corp.

Wiederhold, J. L., & Bryant, B. R. (2001). *Gray oral reading tests-4*. Austin, TX: Pro-Ed.

Windfuhr, K., & Snowling, M. J. (2001). The relationship between paired associate learning and phonological skills in normally developing children. *Journal of Experimental Child Psychology, 80*, 160-173.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3-29.

www.iejee.com

This page is intentionally left blank

International Electronic Journal of Elementary Education, 2011, 4(1), 67-82.

Individual differences in children's knowledge of expository text structures: A review of literature

Melissa N. RAY
The Pennsylvania State University, United States

Bonnie J. F. MEYER *
The Pennsylvania State University, United States

Abstract

In this review of literature we examine empirical research of individual differences in younger readers' knowledge and use of expository text structures. The goal of this review is to explore the influence of reader and text characteristics in order to better understand the instructional needs of elementary school readers. First we review research which has examined the influence of two textual characteristics: the hierarchical organization of macro-and micro-level propositions and the type of text structure (e.g. collection, comparison, problem-and-solution). Then we review research of three reader characteristics: overall comprehension skill, age, and prior knowledge and how their influences may vary in relation to the aforementioned text characteristics. Our review of research suggests that readers of all ages may benefit from explicit instruction in text structure, particularly less-skilled comprehenders. Text structure instruction should focus on highly structured texts like comparison, causation, and problem-and-solution.

Keywords: Text structure, structure strategy, expository text

Introduction

Informational texts can present a challenge for many elementary school readers. In comparison to narrative texts, young readers may have more difficulty comprehending expository texts (Best, Floyd, & McNamara, 2008). This difficulty is likely the result of both the demands on readers' prior knowledge (Best et al., 2008) as well as lower levels of early exposure to expository texts (Duke, 2000). An additional source of difficulty may be the structure of expository texts (Coté, Goldman, & Saul, 1998). In comparison to narrative texts, a larger number of structures are used to describe the organization of expository texts. Meyer (1975, 1985a) proposed five top-level structures: collection, description, comparison,

* Bonnie J. F. Meyer, Department of Educational and Counseling Psychology and Special Education, 204 Cedar Building, Penn State, University Park, PA 16802, United States. Phone: 814-863-7501; FAX: 814-863-1002.
E-mail: bjm8@psu.edu

ISSN:1307-9298

causation, and response (problem-and-solution). When reading expository texts younger readers are confronted unfamiliar concepts that are organized in an unfamiliar way. Difficulties in comprehension of informational texts may lead to trouble with learning from text, particularly in content areas. In order to meet the challenge of reading these texts, readers may employ a number of comprehension processes including: paraphrasing, elaborating (connecting to prior knowledge), and monitoring (Coté et al., 1998).

Another way in which younger readers can successfully comprehend expository text is through their knowledge and use of text structure. Readers who possess a "structure strategy" have knowledge how authors organize texts and seek to apply this knowledge to the text being read, while readers without this strategic knowledge approach text with a "default/list" strategy in which the text is viewed as a collection of loosely related ideas (Meyer, Brandt, & Bluth,1980; see Meyer & Rice, [1982] for a more detailed description of the structure strategy). Readers with structural awareness possess the ability to not only recognize an author's organization, but also to engage in similar organizational processes in establishing their own mental representations of a text.

Awareness of text structure has been associated with better text recall in terms of the number of ideas remembered and their organization (Meyer et al., 1980; Taylor & Samuels, 1983). Structural awareness improves comprehension because it facilitates the construction of a coherent mental representation of text. Coherence, the creation of clear relationships between and among textual ideas contained in one's cognitive representation is considered an essential aspect of text comprehension (van Dijk & Kintsch, 1983; Kintsch, 2004; van den Broek, Young, Tzeng, & Linderholm, 2004). Readers with structural awareness develop a mental representation in which textual ideas are organized according to a defined hierarchical structure. This organization helps readers' ability to remember ideas from a text, especially those propositions at the top of the hierarchy (Britton, Meyer, Hodge, & Glynn, 1980, Meyer et al., 1980; Taylor & Samuels, 1983).

Although some younger readers possess knowledge of expository text structure (Englert & Hiebert, 1984; McGee, 1982; Richgels, McGee, Lomax, & Sheard, 1987; Smith & Hahn, 1987; Yochum, 1991), many do not (Meyer et al., 1980; Taylor, 1980; Taylor & Samuels, 1983). The extent to which children are structurally aware may vary as a function of both textual characteristics such as text structure (Richgels et al., 1987; Smith & Hahn, 1989; Yochum, 1991) as well as reader characteristics, such as age (Englert & Hiebert,1984; Englert & Thomas, 1987; Garner & Gillingham, 1987; McGee, 1982; Smith & Hahn, 1989) and overall comprehension skill (Meyer et al., 1980; Englert & Hiebert, 1984). Intervention research has shown that with explicit instruction in the use of text structure, readers can improve both structural knowledge and comprehension (e.g. Armbruster, Anderson, & Ostertag, 1987; Meyer et al., 2002; Meyer et al., 2010; Williams et al., 2005; Williams et al., 2007; see Meyer & Ray in this issue). However, there is evidence that instruction which is individualized to the instructional needs of the reader is more effective than less individualized instruction (Meyer, Wijekumar, & Lin, 2011). In order to better understand elementary school readers' instructional needs, it is necessary to have a clear understanding of sources of individual differences.

Purpose of this Review of Literature

This review of literature examines sources of individual differences in children's awareness of expository text structures. Previous reviews of literature have described the contribution of reader and text variables on the processing of text structure (e.g. Goldman & Rakestraw, 2000) as well as the interaction between reader and text variables (Roller, 1990). In this review, we investigate how sources of individual differences among readers interact with structural properties to influence both structural knowledge and use of this knowledge in comprehending expository texts. This review differs from Roller (1990) in that we examined the relationship among multiple reader characteristics, while the review conducted by Roller focused specifically on one individual difference: prior knowledge. We take the position that structural awareness varies not only as a function of the reader but also as the reading situation changes. This view reflects current perspectives on reading which stress the importance of the relationship between reader, text, and situational variables (RAND reading study group, 2002). Our goal is to provide a comprehensive picture of students' instructional needs, by providing insights into variability of children's structural knowledge.

Reader and textual characteristics

We used the framework created by Meyer and Rice (1989) to guide the selection of reader and textual influences (see also Meyer, 2003). Meyer and Rice proposed three sources of individual differences in text comprehension: the reader, the text, and the task. Most relevant to the current review are reader and text variables. Reader variables refer to the experiences and skills that readers bring with them to the text and include: age, verbal ability, and prior knowledge (Meyer & Rice, 1989). Text variables refer to those variables which contribute to the content and organization of the text such as text structure, genre, and topic (Meyer & Rice, 1989).

Although there are several possible reader and text influences, we selected those influences which have been both frequently researched and are most relevant for instruction of elementary school readers. With regards to textual characteristics we focus on two major influences. The first is the overall quality of the hierarchical organization and the extent to which subordination of ideas is present. The second source of variability we examined is the type of top-level, rhetorical structure used to organize the text (e.g. comparison, problem-and-solution). We include in this examination of rhetorical structure empirical studies which have examined readers' sensitivity to text signaling. Text signals are textual elements (e.g. headings, overviews, previews, and connectives) which call readers' attention to the structure of the text (Meyer, 1985b; Lorch, 1989). Although this review treats these text characteristics as separate elements, these aspects of text structure are not necessarily mutually exclusive. In this review, we distinguish these elements of structure in order to more closely examine readers' knowledge of various types of top-level-structures as well as their ability to recognize the hierarchical relationships between macro- and micro-level propositions contained in the text. This review does not cover issues of cohesion and its relationship to reader characteristics (see McNamara, Kintsch, Songer & Kintsch,1996; Ozuru, Dempsey, & McNamara, 2009; Voss & Silfies, 1996). Cohesion was excluded from the current review as it covers a broader set of textual characteristics than text structure.

In relation to reader characteristics we focused on three sources of individual differences: overall comprehension skill, age, and prior knowledge. Although each of these variables has been found to influence structural awareness, the nature of this influence is somewhat different. Age and comprehension skill are associated with both the level of awareness a reader has and their ability/need to apply this knowledge. In contrast, prior knowledge is usually associated with the circumstances under which readers do and do not use knowledge of text structure. For each of these reader variables, we explore its unique contribution and how the influence of each of these variables changes in relation to the previous mentioned characteristics of the text.

This review of literature does not represent an exhaustive review of those studies which have examined awareness of text structure; rather we have selected those studies which are representative of research on the influence of reader and textual characteristics. We chose to focus only on those studies which closely align with Meyer's (1975, 1985a) analysis and classification of expository text structures (for alternative approaches to the description and classification of text structure see Alvermann & Hague, 1989; Frederiksen, 1975; van Dijk & Kintsch, 1983). We selected Meyer's approach because several studies of younger readers' structural knowledge as well as interventions designed to teach text structure, reflect this approach. This allowed us to draw meaningful conclusions across studies. Wherever possible, we include those studies which have investigated younger (elementary and middle school) readers.

Influence of Text Characteristics

The overall quality of the hierarchical relationships between propositions within a text may influence readers' ability to perceive and use text structure. The organization of expository text is centered on the hierarchical relationship between macropropositions, higher level ideas, and micropropositions, the lower level ideas which elaborate on them (Weaver & Kintsch, 1991). When the overall organization is compromised, comprehension and recall performance are negatively affected (Kintsch & Yarborough, 1982; Taylor & Samuels,1983; Danner, 1976). Kintsch & Yarborough (1982) found that poorly constructed texts were associated with poorer performance on main idea questions in comparison to well-structured text. Similarly, Danner (1976) and Taylor and Samuels (1983) reported that elementary school readers recalled fewer ideas when reading texts that lacked a clear organization. Taylor and Samuels found the effect of textual organization was related to readers' level of structural awareness. Readers classified as structurally aware had a greater number of ideas recalled with unscrambled over scrambled text, while readers classified as unaware had a similar number of ideas recalled for both scrambled and unscrambled passages (Taylor & Samuels, 1983). This finding suggests that in order for structural knowledge to benefit comprehension, the reader must be able to perceive the hierarchical organization of the text.

Text structure

Readers' structural awareness and the benefits of structural knowledge may also vary in relation to the type of text structure being read. For adults, different text structures may have varying effects on the number of ideas recalled and their organization. Meyer and Freedle (1984) classified the five text structures described by Meyer (1975) on a continuum reflecting the number and ordering of required schematic components contained within each structure. While some structures are less structured, such as collection and description, other structures are more structured, like causation and problem-and-solution. Meyer and Freedle found that more organized structures like comparison and causation were more facilitative of recall (in terms of number of ideas and their organization) than less organized structures like collection. Other researchers have produced similar findings regarding the influence of less and more organized text structures. Sanders and Noordman (2000) and Spooren, Mulder, and Hoeken (1998) found that problem-and-solution was associated with better performance on sentence recognition tasks in comparison to a listing structure (for Spooren et al., 1998 this only occurred with texts in which the structure was marked). Similarly, Wylie and McGuinness (2004) found that the least structured text examined, generalization, was associated with lower recall of main ideas than were the more structured texts, such as comparison.

Previous research on the effect of type of structure on children's knowledge and use text structure has produced mixed findings. Englert and Hiebert (1984) compared 3rd- and 6th-grade readers' sensitivity to description, enumeration, sequence, and comparison text structures using a judgment task in which readers were asked to rate how well a list of sentences fit with the stimulus sentences provided. The researchers found that children had better judgments with the collection structure than the comparison, with performance lowest for comparison and description structures. This difference was present on targets (sentences that fit the structure) but not distractors (sentences that did not fit the structure) (Englert & Hiebert, 1984). Similarly, in their study of 4th -, 6th-, and 8th-graders' recalls, Smith and Hahn (1989) found that comparison was used less frequently to organize recall than enumeration and description. In both of these cases, levels of structural awareness were related to the age of the readers; older readers demonstrated greater knowledge of the comparison structure, than younger readers. This relationship will be further discussed in a later section.

Other researchers have shown patterns of children's awareness that are somewhat similar to the patterns found with adults. Richgels, McGee, Lomax, & Sheard (1987) found that 6th-grade readers were more sensitive to the comparison structure than the other structures examined (collection, causation, problem-and-solution). They found that causation texts posed the greatest challenge to 6th-grade readers; performances on a matching task as well as organizational ratings of recalls and compositions were lowest for causation (Richgels et al., 1987). Similarly, Yochum (1991) found that 5th graders recalled more ideas when reading comparison texts than when reading attribution (collection) texts. However, this effect for text structure was not found on a comprehension test. These findings suggest that like adults, children may benefit from more organized text structures. However, unlike adults, children's knowledge of text structure may still be developing, and those structures that are most structured like the causation structure, may pose a challenge.

In addition to the overall rhetorical structure, textual signaling devices that explicitly indicate the structure of the text may also influence readers' abilities to recognize and use the text structure. Numerous studies which have examined the influence of text signaling on adult comprehension have demonstrated that signaling is associated with a greater number of ideas recalled and more organized recalls (Kardash & Noel, 2000; Lorch & Lorch, 1985; Lorch, Lorch, & Inman, 1993; Ritchey, Schuster, & Allen, 2008; for an extensive review of text signaling see Lorch, 1989). Research on younger readers' sensitivity to signaling reveals similar effects. Ohlhausen and Roller (1988) in their examination of younger readers' sensitivity to structural and content schemas found that texts which contained elements highlighting both the structural and content organization were associated with a greater number of correct main ideas identified in comparison to those texts which emphasized either the structure or the content. Moreover readers exposed to a text which emphasized the text structure were better able to identify the structural organization (Ohlhausen & Roller, 1988). Similarly, Rossi (1990) found that structural schemas which highlighted macrosentences and/or the overall rhetorical structure were associated with higher performance on a test of comprehension and a larger number of text propositions produced in 5^{th}-grade readers' summaries. The effectiveness of signals has been found to vary in relation to the overall comprehension skill of the reader (Meyer et al., 1980). Nevertheless, previous research suggests that providing indices of the top-level structure of texts can facilitate readers' structural awareness by making the structure more explicit.

The Relationship between Reader and Text Characteristics

In this section, we will examine previous research which has investigated sources of individual differences (reading ability, age, prior knowledge) in structural awareness and their relationship to text characteristics. Although we will address these sources individually, many studies consider multiple reader and text influences simultaneously. Table 1 contains a description of studies which have examined individual differences in younger readers' knowledge of text structure.

Table 1. *Reader and Text Influences on Structural Knowledge*

Authors	*Date*	*Reader*	*Text*	*Text Structure(s)*
Danner	1976	Age	HQ	Description
Taylor	1980	Age, CA		Attribution
Meyer et al.	1980	CA, SA	S	Comparison, problem-solution
McGee	1982	Age, CA		Description
Taylor & Samuels	1983	SA	HQ	
Loman & Mayer	1983	CA	S	
Englert & Hiebert	1984	Age, CA	TS	Description, enumeration, sequence, comparison
Garner et al.	1986	Age	HQ	
Garner & Gillingham	1987	Age	HQ	
Englert & Thomas	1987	Age, LD	TS	Description, enumeration, sequence, comparison
Richgels et al.	1987	none	TS	Collection, comparison causation, problem-solution
Ohlhausen & Roller	1988	Age	S	Description
Smith & Hahn	1989	Age	TS	Description, enumeration sequence, and comparison
Rossi	1990	CA	S	Problem-solution
Yochum	1991	PK	TS	Attribution, comparison
Vauras et al.	1994	Age, CA		Description, functional/causal
Armand	2001	PK	TS	Causation+collection/ casuation+comparison
Ray	2011	Age, CA	TS	Comparison, problem-solution

Note. Reader variables: CA= comprehension ability, LD= learning disability, PK= prior knowledge, SA= structural awareness level; Text variables: HQ = hierarchical quality, S= signaling, TS= text structure

Reading comprehension skill

Readers' overall comprehension abilities are associated with their levels of structural awareness. In general, previous research has found that readers classified as skilled comprehenders have greater structural awareness in comparison to less skilled readers (Englert & Hiebert, 1984; Hiebert, Englert, & Brennan, 1983; McGee, 1982; Meyer et al., 1980; Ray, 2011; Taylor, 1980).Within the same grade, skilled comprehenders recall more ideas after reading and produce recalls that more closely reflect the author's organization (McGee, 1982; Meyer et al., 1980; Ray, 2011; Taylor, 1980). In addition, Englert and Hiebert (1984) found that on rating tasks, highly skilled readers were better able to identify statements that matched the structure of the stimulus sentences and to exclude those that did not.

These findings suggest that skilled readers are more likely to have knowledge of text structure and to approach text with a "structure strategy", and as a result, they are more likely to establish a well-organized mental representation of the text. This organized mental representation in turn improves recall. However, the benefit on recall may be restricted to those ideas located at the topic of the hierarchical structure with little difference among comprehension skill groups on recall of details (McGee, 1982). This difference may reflect structural awareness, as text structure is likely most facilitative in retrieval of macro-level propositions (Britton et al., 1980). Although overall comprehension ability has been associated with greater structural awareness, the nature of this relationship remains unclear. It not clear whether greater comprehension skill causes higher levels of structural knowledge, or whether structural knowledge contributes to improvements in overall comprehension skill. Intervention research (e.g., Meyer et al., 2010) does indicate that instruction of text structure is associated with improved performance on standardized tests of reading comprehension. Thus, it is possible that structural knowledge can predict as well as affect readers' comprehension abilities.

While the relationship between comprehension skill and structural awareness appears to be consistent across text structures (Englert & Hiebert, 1984; Meyer et al., 1980), the effect of text signals may vary in relation to the overall comprehension skill of the reader. While some studies have found that both skilled and less killed readers benefit from explicit indicators of text structure (Loman & Mayer, 1983; Rossi, 1990) others have not (Meyer et al., 1980). Meyer et al. (1980) found that text signaling benefited immediate recall for readers classified as "underachievers" (readers for whom a discrepancy existed between word reading and text comprehension) but did not have an effect on readers classified as high and low comprehenders. This benefit for signaling was found for only one of the two passages students read, a problem-and-solution text (Meyer et al., 1981). There was no effect for signaling on the comparison text read.

Although Rossi (1990) found that reading text with signaling helped both skilled and less skilled readers in comparison to reading texts with no signaling, skilled and less skilled readers differed with respect to which types of signaling were most beneficial. Rossi examined the effect of multiple signaling conditions in which the presence of two types of signaling, underlining of macrostructure sentences (topic sentences) and headings which indicated the text structure, were manipulated. Skilled readers performed better with signaling only in those signaling conditions in which the macro-sentences were underlined, while less skilled readers performed similarly across signaling conditions (Rossi, 1990).

The relationship between signaling and reading ability is complex. Although text signaling may help readers to apply their knowledge of structure to a particular text, the effectiveness of these signals is relative to the readers' need and ability to make use of them. Highly skilled readers may not need text signaling (or certain text signals) in to order apply structural knowledge, while readers with particularly low comprehension skills may have such low levels of structural knowledge that they cannot take advantage of these signaling devices.

Age

Knowledge of expository text structures and structural properties of texts may increase with age as children gain more exposure to expository texts. Studies which have compared differences in age groups' awareness of the various expository text structures have found that levels of awareness increased with age (Danner, 1976; Englert & Hiebert, 1984; Englert & Thomas, 1987; McGee, 1982; Ray, 2011; Smith & Hahn, 1989; Taylor, 1980). In general, older readers recall more ideas after reading (Danner, 1976; McGee, 1982; Ray, 2011; Taylor, 1980), produce recalls which more closely reflect the author's organization (Danner, 1976; McGee, 1982; Ray, 2011; Smith & Hahn, 1989; Taylor, 1980), and perform better on rating tasks measuring the identification of text structure (Englert & Hiebert, 1984; Englert & Thomas, 1987). This increase with age may be related to the relative ages of the readers. Ray (2011) compared 4th-, 6th-, 7th- and 9th-grade readers' awareness of problem-and-solution and comparison texts using written recalls (problem-and-solution) and main idea statements (comparison). Although structural awareness improved with age (particularly from 4th to 6th grades), fewer significant differences in performance were found among older readers (grades 6-9), suggesting that while structural awareness increases with age, there may be periods where little change occurs. Lack of change was not a result of mastery of knowledge about text structure in the higher grades because similar to Meyer et al. (1980) about 50% of the 9th-graders gave evidence for good use of the problem-and-solution text structure.

Age may also predict readers' knowledge of other structural properties of expository texts. Both younger and older readers have knowledge of how texts should be organized, but older readers may have a more sophisticated and developed sense of organization. Danner (1976) found that even with prompting few 2nd-graders could attribute differences between ill-structured and well-structured passages to the text structure, while with prompting 4th- and 6th-graders could describe the organizational differences. Similarly, Ohlhausen and Roller (1988) found that 9th-graders were more likely than 5th-graders to provide structural reasons for the selection of main ideas in an underlining task.

In their study of 3rd-, 5th-, and 7th-grader's knowledge of structure, Garner et al., (1986) asked readers to complete a paragraph recognition task (identify a paragraph) and a paragraph construction task in which readers were given a set of related and unrelated sentences. Although all readers could identify segments of text as paragraphs, construct paragraphs which included topically related sentences, and place topic sentences appropriately, there were age differences in some areas of paragraph construction. Seventh grade readers were better at excluding unrelated sentences and establishing cohesion by organizing related sentence pairs to be adjacent. In a similar study, Garner and Gillingham (1987) asked 5th and 7th graders to compose a good and bad paragraph using a set of given sentences and to provide a verbal report of their decisions. These sentences included intrusions (sentences unrelated to the topic). Seventh graders scored higher on verbal reports of topic relatedness and superordination (placement of topic sentence) (Garner & Gillingham, 1987). However, age differences in performance on topic relatedness and superordination measures for good paragraphs showed no significant differences between the groups (Garner & Gillingham, 1987). Although younger readers used superordination and topic grouping to organize paragraphs, they were less likely to report it.

Age and text structures

In addition to awareness of the hierarchical relationships between ideas, knowledge of various types of text structures may increase with age. Englert and Hiebert (1984) found an interaction between grade and text structure on distractor items (ability to determine that a sentence did not fit the structure of the stimulus sentences). For 6th-graders there was no difference in performance across text structures, while 3rd-graders had lower ratings of distractors for sequence and description in relation to comparison (Englert & Hiebert, 1984). For 3rd-grade readers, their ability to detect that a sentence did not belong varied as function of the text structure, but for 6th graders, it did not. Smith and Hahn (1989) compared readers in grades 4, 6, and 8 on oral recalls of enumeration, comparison, and, sequence and found that all readers used the enumeration and description structures to organize their oral recalls. In contrast, few 4th and 6th graders used the comparison text structure after reading comparison texts, while many 8th-grade readers did (Smith & Hahn, 1989). Overall, previous research suggests that as readers age, their knowledge moves across the continuum that Meyer and Freedle (1984) proposed with students gaining knowledge of more organized structures like comparison. These more organized structures may provide more knowledge hooks to improve processing and memory for readers who can use the structures (Meyer & Freedle).

Age and comprehension skill

Structural awareness appears to increase as children age; however, previous research also indicates that the development of structural awareness may be constrained by the overall comprehension ability of the reader. Taylor (1980) examined the recalls of 6th-grade skilled and less skilled readers as well as 4th graders. At immediate recall, 6th-grade readers recalled more ideas than 4th graders, with no difference between reading skill groups. However, at delayed recall, skilled 6th-grade readers recalled more ideas than less skilled readers and 4th-grade readers, with no significant difference between the latter groups (Taylor, 1980). Readers performed similarly on measures of organization of recall. At immediate post-test there was no difference between grades, while at delayed recall, more skilled 6th-grade readers organized their recalls according to the structure of the text in comparison to the other groups, with no significant difference between 4th graders and less skilled 6th graders (Taylor, 1980). At delayed recall only those 6th-grade readers classified as skilled demonstrated higher levels of structural awareness in comparison to the 4th-grade group.

Research on the development of text processing, has also revealed that increases with age may be related to the overall comprehension skill of the reader. In a longitudinal study, Vauras, Kinnunen, & Kuusela (1994) compared changes in text processing strategies in grades 3 through 5 for readers classified as high, middle, and low. The researchers found that all readers improved in their organization of recalls according to the text structure, but for low readers these differences were not statistically significant (Vauras et al., 1994). High and middle groups of readers improved in both local (e.g., integrating ideas found in adjacent sentences) and global coherence (e.g., integrating across paragraphs), but low readers significantly improved only in local coherence (Varuas et al., 1984). Moreover, in comparing high and low readers, Vauras et al. found that high readers made larger gains in structural processing than low readers. Taken together, these findings suggest that development of readers' structural awareness is related to their overall reading ability, while skilled readers acquire the ability to approach text with a structure strategy, less skilled readers may continue to approach text with a default/list strategy.

Other research suggests that while reading skill may predict differences within grades, it does not predict difference across grades. McGee (1982) compared recalls of skilled and less skilled 5th- and skilled 3rd-graders. McGee found that both high and low skilled 5th-grade readers recalled more ideas than 3rd-grade readers. The structure of recalls also improved across groups, with skilled 5th-graders demonstrating use of the authors' structure, less skilled readers showing partial use, and 3rd- graders showing no use of the author's structure (McGee, 1982). Similarly, Englert & Hiebert (1984) and Ray (2011) failed to find a significant interaction between reading ability and age. However, Ray (2011) did find that when the overall comprehension ability of the grades compared, in this case 6th- and 7th-grade, were not significantly different, no significant difference in performance on structural awareness measures was found, suggesting that increases in structural knowledge are related to the overall comprehension skill of the reader. Nevertheless, these findings suggest that structural awareness increases with age for both skilled and less skilled readers. Findings from the longitudinal study by Vauras et al. (1994) suggest the need for more longitudinal studies focusing on growth in structural awareness across elementary and middle school years comparing below grade-level vs. grade level and above readers.

Prior knowledge

In comparison to reading ability and age, relatively few studies have investigated the influence on use of text structure of younger readers' prior knowledge about the topic domain of a text. Readers' prior knowledge of text content may influence the extent to which readers use text structure and benefit from indicators of text structure. Research with adults suggests that prior knowledge influences processing of the hierarchical structure of expository texts (Birkmire, 1985) and the relative benefits of text signaling (Lorch & Lorch, 1996). In relation to text signaling, Lorch and Lorch (1996) found that signaling was less effective when text topics were familiar to readers, than when they were unfamiliar. This finding suggests that readers who have prior knowledge may be less reliant on cues to the text structure. In their study of prior knowledge and text structure, Wylie and McGuinness (2004) found that both low and high prior knowledge readers benefited from reading texts with more structure (comparison, sequence, classification, and enumeration) than for the least structured texts (generalization). When comparing relative performance of these more structured texts, Wylie and McGuinness found that high prior knowledge readers recalled more ideas with comparison, but there were no differences in recall across these four text structures for low prior knowledge readers. Although both groups used structural knowledge, when prior domain knowledge was low readers may have been less successful in applying structural knowledge about texts. Overall research of adults' text processing suggests that readers' application of structural knowledge is related to their prior content knowledge.

Research findings regarding the relationship between children's prior domain knowledge and text structure have been mixed. Armand (2001) examined the relationship between French 6th-grade readers' prior topic knowledge and text structure in their influence on text comprehension. The author compared the performance of readers with high and low prior knowledge on the comprehension of texts with a combined text structure of causation plus comparison (causative agent acid rain affecting numerous attributes [i.e., bark] of two kinds of trees discussed attribute by attribute with comparative signaling comparing each tree for each attribute) versus causation plus collection (causative agent acid rain affecting two types of trees with the trees discussed one at a time going through each tree's attributes before discussing the second tree and its attributes). The second presentation method was thought to be less complex and less demanding on processing. The findings were complex because they varied with the classification of questions (e.g, recall [open ended] vs. recognition [closed multiple choice question]). The open-ended questions showed that for high knowledge students, the causation plus comparison yielded better results than causation plus collection, while the opposite was true for low knowledge students. Armand also found a significant interaction between prior knowledge and text structure for the multiple-choice questions. However, for high knowledge readers there was no difference in performance across text structures, whereas low prior knowledge readers had better performance with the causation plus collection structure. These findings suggest that prior knowledge mitigates the demands of reading texts containing two, challenging yet more organized, structures for younger readers. Yekovich, Walker, Ogle, and Thompson (1990) found that high domain knowledge (i.e., football) enabled low aptitude students to handle causal relationships and other types of inferences above what was expected based on their aptitude test scores. In the Armand study high prior knowledge enabled 6th graders to process ideas from text in the more complex causal and comparative text, while those with low prior knowledge found the complex causal and comparative text too difficult. Students with low prior knowledge could better understand the less processing intense causation plus collection with its attributes grouped in a collection related to each tree. Similar to Wylie and McGuinness (2004), Armand's finding also suggests that low prior knowledge readers may be less able apply structural knowledge when the content demands are high. The causal structure had been found to be particularly difficult with 6th graders (Richgels et al., 1987). In the Armand study with high prior knowledge, text structure differences did not matter for recognition, but only with the deeper processing needed for recall questions. Some text structures may be unfamiliar, or too difficult for a certain grade levels or knowledge or skill levels within a grade level. Both the Armand as well as the Wylie and McGuinness (2004) study discussed above suggest that with high domain knowledge, the comparison structure can be helpful for deeper understanding.

Unlike the more complex texts studied by Armand (2001), Yochum (1991) examined the effect of attribution (collection of descriptions) vs. comparison structures on 5th-graders' recall and performance on comprehension questions. Yochum failed to find an interaction between text structure and prior knowledge. Both prior knowledge and text structure predicted the number of ideas recalled, while only prior knowledge predicted comprehension test scores. Neither prior knowledge nor text structure was associated with the structure of recalls. This finding suggests that while both prior knowledge and text structure influence comprehension, prior knowledge does not moderate the effect of text structure. However, one should be cautious about drawing this conclusion because the structure of the text did not influence organization of recall. Previous studies have indicated that readers' of this age have varying levels of sensitivity to these structures (e.g. Richgels, et al. 1987).

Few studies and inconsistencies in findings make it difficult to draw substantive conclusions about the relationship between prior knowledge about text content and/or text structure and measures of structural awareness. This probably relates to the complexity of the interactions between the reader variables of domain knowledge, structure strategy knowledge with different types structures, general reading skills, and the particular text to be read as well as the task demands. Nevertheless prior knowledge may impact readers' need to avail themselves to the affordances of text structures. Additionally, the ability to apply knowledge of text structure may vary with the type of text structure used in a text (e.g., causal vs. comparative structures).

Summary of Research Findings

Previous research indicates that the interaction between the knowledge and skills of the reader and the structural characteristics of the text influences readers' structural awareness. Older readers possess knowledge of a larger set of text structures in comparison to younger readers. In developing awareness of expository text structures, readers' move from a list like/default strategy to a "structure strategy" as proposed by Meyer et al. (1980). This development of structural knowledge reflects the continuum of structural organization proposed by Meyer and Freedle (1984). Less organized structures are likely acquired before more organized and more organized structures. Previous research indicates that by 6th-grade readers demonstrate awareness of a variety of text structures (Richgels et al., 1987, Englert & Hiebert, 1984). Development of structural awareness may differ for less skilled readers, as lower overall comprehension skills may hinder these readers' abilities to acquire greater levels of structural awareness. However, additional research is needed to further explore the influence comprehension skill has on the development of structural awareness. Moreover, even if readers have structural knowledge, the content knowledge demands of a text may inhibit their ability to apply structural knowledge. Alternatively, highly familiar domains for readers may make the affordances of text structure unnecessary. Ultimately readers' ability to perceive and use the authors' structure is the result of their ability to meet both the structural and the content demands of a text.

Although many studies have examined sources of individual differences, few studies have examined multiple sources simultaneously (See Table 1). In particular, few studies have examined the influence of comprehension skill on multiple text structure structures (Englert & Hiebert, 1984; Meyer et al., 1980). Although several studies examined age differences, few studies have examined the relationship between age and comprehension skill (e.g. Taylor, 1980; McGee, 1982; Vauras et al., 1994). Only one study in this review examined the influence of age and comprehension ability on awareness of several text structures (Englert & Hiebert, 1984). Moreover, few studies have examined the influence of younger readers' prior knowledge on the use of expository texts structures (Armand, 2001; Yochum, 1991). Even though several studies have examined multiple text structures, few of these included the causation and problem-and-solution structures. None of the studies reviewed have examined the relationship between all of the sources of individual differences (age, reading ability, and prior knowledge). Additional research is needed to explore these complex relationships, using multiple text structures, particularly more organized text structures like causation and problem-and-solution as these may be more sensitive to changes in reader characteristics. A closer examination of how sources of individual differences vary across reading situations or time in longitudinal designs will provide a clearer picture of students' needs.

Instructional Implications

Most elementary school readers would likely benefit from explicit instruction in the use of expository text structures. All students, but particularly younger readers, need instruction in more organized text structures like comparison, causation, and problem-and-solution. Not only do these structures provide maximal benefits for memory of expository texts, these structures may also pose the greatest challenge to younger readers. Despite age increases in structural awareness, previous research indicates that later elementary and middle school students continue to find these structures difficult (Richgels et al., 1987; Meyer et al., 1980). Students in elementary grades may also benefit from instruction in the hierarchical structure of expository texts. Although this review did not include writing research, it appears from research which has used text construction (Garner et al., 1986; Garner & Gillingham, 1987), that younger readers may need more practice in writing expository texts.

In addition, less skilled readers are likely in the greatest need of intensive and explicit instruction in text structure, including instruction in text signaling. In an intervention study, Meyer et al. (2010) found that an interaction between type of structure strategy instruction (elaborated feedback with scaffolding vs. just giving information about correctness of answers) and reading ability (below grade level vs. grade level or above) predicted 31% of the variance in who would make large gains in competency using the problem-and-solution structure before and after instruction. Below-grade-level readers were more likely to jump from no awareness of the problem-and-solution structure to competency using the problem-and-solution structure if they received elaborated feedback with modeled responses. However, the feedback condition did not make a difference for better readers. Poorer readers improved only with elaborated feedback, while better readers improved with structure strategy instruction with both types of feedback. In general, previous research which has examined development of structural awareness, suggests that these students may fail to develop competency in structural awareness without explicit instruction. These readers may also benefit from instruction in text signaling as they may be less sensitive to these devices.

Teachers need to carefully select texts when teaching readers how recognize and use expository text structure. Texts which are well organized and clearly reflect the structure being taught will help readers to apply structural knowledge. Teachers should also consider readers' prior knowledge of the text topics. Texts which contain difficult and unfamiliar material may pose a challenge to readers' ability to apply structural knowledge. However, once readers gain greater levels of structural awareness, they should also be provided opportunities to read a variety of texts. Finally, because readers vary in their knowledge and ability to use expository text structures, it is important for teachers to assess reader's levels of structural awareness prior to instruction. Awareness studies have relied heavily on recalls, usually written. While these are easy measures to administer they may be challenging for teachers to analyze and interpret. Additional research is needed on classroom-based assessments of children's structural awareness.

Conclusion

Knowledge of text structure may help younger reader to overcome the difficulties of reading expository texts. However, in order to gain competency in recognizing and using expository text structure, many students will likely require explicit instruction. In designing effective, text structure instruction, it is important to understand the needs of individual readers. Unfortunately, the individual needs of a particular reader may not always be clear given the complex nature of the relationship between reader and text. Additional research which examines the simultaneous contributions of a variety of influences may be able to provide much needed insights into meeting the needs of diverse groups of learners.

• • •

Melissa N. Ray is a doctoral candidate in Educational Psychology at The Pennsylvania State University. Her primary research interests are reading comprehension, literacy development, and literacy instruction.

Bonnie J. F. Meyer, Professor of Educational Psychology at The Pennsylvania State University, studies reading comprehension, decision-making, text structure, and the structure strategy. She has over 50 publications about the structure strategy and its effects on understanding and applications for reading or writing. Her current research, funded by the Institute for Education Sciences of the U.S.

Department of Education, is focused on structure strategy instruction for students in Grades 4-8. She has served on editorial boards for *Discourse Processes*, the *Journal of Educational Psychology*, the *Reading Research Quarterly*, *Cognition and Instruction*, the *Educational Psychologist*, and the *Journal of Literacy Research* and is a Fellow in the American Psychological Association, American Educational Research Association, and American Psychological Society.

References

Alvermann, D. E., & Hague, S. A. (1989). Comprehension of counterintuitive science text: Effects of prior knowledge and text structure. *The Journal of Educational Research, 82*, 197-202.

Armand, F. (2001). Learning from expository texts: Effects of the interaction of prior knowledge and text structure on responses to different question types. *European Journal of Psychology of Education, 16*, 67-86.

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.

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29*, 137-164.

Birkmire, D. P. (1985). Text processing: The influence of text structure, background knowledge, and purpose. *Reading Research Quarterly, 20*, 314-326.

Britton, B. K., Meyer, B. J., Hodge, M. H., & Glynn, S. M. (1980). Effects of the organization of text on memory: Tests of retrieval and response criterion hypotheses. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 620-629.

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.

Danner, F.W. (1976). Children's understanding of intersentence organization in the recall of short descriptive passages. *Journal of Educational Psychology, 68,* 174-183.

Duke, N. K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly, 35,* 202-224.

Englert, C. S., & Hiebert, E. H. (1984). Children's developing awareness of text structures in expository materials. *Journal of Educational Psychology, 76*, 65-74.

Englert, C. S., & Thomas, C. C. (1987). Sensitivity to text structure in reading and writing: A comparison between learning disabled and non-learning disabled students. *Learning Disability Quarterly, 10*, 93-105.

Frederiksen, C. H. (1975). Representing logical and semantic structures of knowledge acquired from discourse. *Cognitive Psychology, 7,* 371-458

Garner, R., & Gillingham, M. G. (1987). Students' knowledge of text structure. *Journal of Reading Behavior, 19*, 247-259.

Garner, R., Alexander, P., Slater, W., Hare, V. C., Smith, T., Reis, R. (1986). Children's knowledge of structural properties of expository text. *Journal of Educational Psychology, 78*, 411-416.

Goldman, S. R., Rakestraw J. A. (2000). Structural aspects of constructing meaning from text. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (Eds.), *Handbook of reading research: Volume III* (pp. 311-335). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hiebert, E. H., Englert, C. S., & Brennan, S. (1983). Awareness of text structure in recognition and production of expository discourse. *Journal of Reading Behavior, 15*, 63-79.

Kardash, C. M., & Noel, L.K. (2000). How organizational signals, need for cognition, and verbal ability affect text recall and recognition. *Contemporary Educational Psychology, 25,* 317-331.

Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. B. Ruddell, & U. J. Norman (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1270-1327). Newark, DE: International Reading Association.

Kintsch, W., & Yarbrough, J. C. (1982). Role of rhetorical structure in text comprehension. *Journal of Educational Psychology, 74*, 828-834.

Loman, N. L., & Mayer, R. E. (1983). Signaling techniques that increase the understandability of expository prose. *Journal of Educational Psychology, 75*, 402- 412.

Lorch, R.F. (1989). Text-signaling devices and their effects reading and memory processes. *Educational Psychology Review, 1,* 209-234.

Lorch, R. F., & Lorch, E. P. (1985). Topic structure representation and text recall. *Journal of Educational Psychology, 77*, 137-148.

Lorch, R. F., Lorch, E. P., & Inman, W. E. (1993). Effects of signaling topic structure on text recall. *Journal of Educational Psychology*, 85, 281-290.

Lorch, R.F., & Lorch, E.P. (1996). Effects of headings on text recall and summarization. *Contemporary Educational Psychology, 21,* 261-178.

McGee, L. M. (1982). Awareness of text structure: Effects on children's recall of expository text. *Reading Research Quarterly, 17,* 581-590.

McNamara, D.S., Kintsch, E. Songer, N.B., & 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.

Meyer, B. J. F. (1975). *The organization of prose and its effects on memory*. Amsterdam: North Holland.

Meyer, B. J. F. (1985a). Prose analysis: Purposes, procedures, and problems. In B. K. Britton, & J. Black (Eds.), *Analyzing and understanding expository text* (pp. 11-64, 269 304). Hillsdale, NJ: Erlbaum.

Meyer, B. J. F. (1985b). Signaling the structure of text. In D.H. Jonassen (Ed.), *The technology of text* (pp. 64–89). Englewood Cliffs, NJ: Educational Technology.

Meyer, B. J. F. (2003). Text coherence and readability. *Topics in Language Disorders, 23,* 204-224.

Meyer, B. J. F., Brandt, B. J., & 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., & Freedle, R. O. (1984). Effects of discourse type on recall. *American Educational Research Journal, 21*, 121-143.

Meyer, B. J. F., Middlemiss, W., Theodorou, E., Brezinski, K. L., McDougall, J., & Bartlett, B. J. (2002). Effects of structure strategy instruction delivered to fifth grade children using the internet with and without the aid of older adult tutors. *Journal of Educational Psychology, 94,* 486-519.

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., Wijekumar, K., Middlemiss, W., Higley, K., Lei, P., Meier, C., & Spielvogel, J. (2010). Web-based tutoring of the structure strategy with or without elaborated feedback or choice for fifth- and seventh-grade readers. *Reading Research Quarterly, 45,* 62-92.

Meyer, B. J. F., Wijekumar, K., & Lin, Y. -C. (2011). Does individualization of web-based structure strategy instruction increase fifth graders' reading comprehension. *Journal of Educational Psychology, 103,* 140-168

Ohlhausen, M. M., & Roller, C. M. (1988). The operation of text structure and content schemata in isolation and in interaction. *Reading Research Quarterly, 23*, 70-88.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction, 19,* 228-242.

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND. Retrieved April 27, 2009 from http://www.rand.org/multi/achievementforall/reading/readreport.html

Ray, M.N. (2011). *The development of structural awareness: The relationship between age, skill, and text* (Unpublished Master's Thesis). The Pennsylvania State University, University Park.

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.

Ritchey, K., Schuster, J., & Allen, J. (2008). How the relationship between text and headings influences readers' memory. *Contemporary Educational Psychology, 33,* 859–874.

Roller, C. M. (1990). The interaction between knowledge and structure variables in the processing of expository prose. *Reading Research Quarterly, 25*, 79-89.

Rossi, J. P. (1990). The function of frame in the comprehension of scientific text. *Journal of Educational Psychology, 82*, 727-732.

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.

Smith, T. F., & Hahn, A. L. (1989). Intermediate-grade students' sensitivity to macrostructure intrusions. *Journal of Reading Behavior, 21*, 167-180.

Spooren, W., Mulder, M., & Hoeken, H. (1998). The role of interest and text structure in professional reading. *Journal of Research in Reading, 21*, 109-120.

Taylor, B.M. (1980). Children's memory for expository text after reading. *Reading Research Quarterly, 15,* 399-411.

Taylor, B. M., & Samuels, S. J. (1983). Children's use of text structure in the recall of expository material. *American Educational Research Journal, 20*, 517-528.

Vauras, M., Kinnunen, R., & Kuusela, L. (1994). Development of text-processing skills in high-, average-, and low-achieving primary school children. *Journal of Reading Behavior, 26*, 361-389.

Van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (2004). The landscape model of reading: Inferences and the online construction of a memory representation. In R. B. Ruddell, & U. J. Norman (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1249-1269). Newark, DE: International Reading Association.

Voss, J. F., & Silfies, L. N. (1996). Learning from history text: The interaction of knowledge and comprehension skill with text structure. *Cognition and Instruction, 14*, 45-68.

Weaver, C. A. I., & Kintsch, W. (1991). Expository text. In R. Barr, M. L. Kamil, P. B. Mosenthal & P. D. Pearson (Eds.), *Handbook of reading research volume II* (pp. 230-245). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

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.

Williams, J. P., Nubla-Kung, A. M., Pollini, S., Stafford, K. B., Garcia, A., & Snyder, A. E. (2007). Teaching cause-effect text structure through social studies content to at-risk second graders. *Journal of Learning Disabilities. Special Issue: Teaching History--in all its Splendid Messiness--to Students with LD: Contemporary Research, 40,* 111-120.

Wylie, J., & McGuinness, C. (2004). The interactive effects of prior knowledge and text structure on memory for cognitive psychology texts. *British Journal of Educational Psychology, 74*, 497-514.

Yekovich, F. R., Walker, C. H., Ogle, L. T., & Thompson, M. A. (1990). The influence of domain knowledge on inferencing in low-aptitude individuals. *The Psychology of Learning and Motivation, 25*, 259-278.

Yochum, N. (1991). Children's learning from informational text: The relationship between prior knowledge and text structure. *Journal of Reading Behavior, 23*, 87-108.

International Electronic Journal of Elementary Education, 2011, 4(1), 83-106.

The differential relations between verbal, numerical and spatial working memory abilities and children's reading comprehension

Jane OAKHILL *
School of Psychology University of Sussex, Brighton, United Kingdom

Nicola YUILL
School of Psychology University of Sussex, Brighton, United Kingdom

Alan GARNHAM
School of Psychology University of Sussex, Brighton, United Kingdom

Abstract

Working memory predicts children's reading comprehension but it is not clear whether this relation is due to a modality-specific or general working memory. This study, which investigated the relations between children's reading skills and working memory (WM) abilities in 3 modalities, extends previous work by including measures of both reading comprehension and reading accuracy. Tests of word reading accuracy and reading comprehension, and working memory tests in three different modalities (verbal, numerical and spatial), were given to 197 6- to 11-year old children. The results support the view that working memory tasks that require the processing and recall of symbolic information (words and numbers) are better predictors of reading comprehension than tasks that require visuo-spatial storage and processing. The different measures of verbal and numerical working memory were not equally good predictors of reading comprehension, but their predictive power depended on neither the word vs. numerical contrast nor the complexity of the processing component. In general, performance on the verbal and numerical working memory tasks predicted reading comprehension, but not reading accuracy, and spatial WM did not predict either. The patterns of relations between the measures of working memory and reading comprehension ability were relatively constant across the age group tested.

Keywords: Reading comprehension, reading accuracy, working memory, information processing

Introduction

The concept of working memory has a role in most theories of text comprehension, and in attempts to explain individual differences in text comprehension (see, e.g. Just and Carpenter, 1992). Daneman and Carpenter (1980) suggested that the crucial difference

* Jane Oakhill, School of Psychology, Pevensey 1 Building, University of Sussex, Falmer, BN1 9QH, UK.
E-mail: j.oakhill@sussex.ac.uk

ISSN:1307-9298

www.iejee.com

between tests of working memory and those of short-term memory (such as digit span and word span) which are not, or are only weakly, related to comprehension skill, is that short-term memory tests only require the use of a passive storage buffer. Daneman and Carpenter went on to argue that both storage and processing of information in memory is important in comprehension, and suggested that the concept of working memory (e.g. Baddeley & Hitch, 1974) better accounts for the sharing of resources between the processing and storage demands of a particular task. In order to measure this functional capacity, Daneman and Carpenter developed the *Reading Span* task. In contrast to digit span and related tasks, performance on both reading and listening versions of Daneman and Carpenter's working memory span tasks predicted performance on comprehension tests.

Daneman and Carpenter's (1980; 1983) reading span test is now a frequently used measure of working memory in reading research. In this test, participants either read or listen to a set of unrelated sentences (processing requirement) and have to retain the final word of each sentence (storage requirement) for recall after all the sentences have been read. Participants also have to answer simple comprehension questions about the sentences to ensure that they have processed the text for meaning. Studies of college students have shown that scores on this test correlate highly with many measures of reading comprehension such as remembering facts, detecting and recovering from semantic inconsistencies, and resolving pronouns, especially those with distant antecedents (see, e.g., Cantor, Engle, & Hamilton, 1991; Dixon, LeFevre, & Twilley, 1988; Engle, Nations, & Cantor, 1990). The correlations of span with performance on various comprehension tests ranged from .7 to .9 in the original samples, and a meta-analysis by Daneman and Merikle (1996), shows an average correlation of .41 between reading span and global reading comprehension in adults.

The link between working memory and reading comprehension probably holds because a major component of skilled comprehension is the ability to compute the semantic and syntactic relations among successive words, phrases and sentences, in order to construct a coherent overall representation of the text. In all current models of text comprehension (e.g. Gernsbacher, 1990; Johnson-Laird, 1983; Kintsch, 1998) the processes of integration and inference are important in the construction of a coherent model of the text, both locally and globally. In such models, working memory acts as a buffer for the most recently read propositions in a text, so that they can be integrated with the model of the text so far, and also holds information activated from long term-memory to facilitate its integration with the currently active text (Cooke, Halleran & O'Brien, 1998; Graesser, Singer & Trabasso, 1994). It follows that individuals with limited working memory capacities should be less able to undertake these types of processing than those with greater storage and processing capacities. However, Daneman and Carpenter's original reading span test itself requires reading, so performance on the test may be partly, or even largely, dependent on general reading ability, which is known to be correlated with reading comprehension skill, but which involves many components other than working memory. Furthermore, people have to perform comprehension tasks, albeit simple ones, as an integral part of the reading span task. Such considerations raise the question of what underlies the relation between reading span and reading comprehension. Thus, given more general arguments for the role of working memory in comprehension, the question remains open as to whether the type of working memory implicated in text comprehension is a general one, one that is specific to language, or to the processing of symbolic information. On the one hand, several studies support the idea that it is the processing of symbolic information that is crucial. These studies show that verbal and numerical span tasks, but not spatial span tasks, predict performance on tests of reading comprehension and other measures of verbal ability (see, e.g. Daneman & Tardif, 1987; Shah & Miyake, 1996) whereas spatial span, but not reading span, is a good predictor of performance on standardised visuo-spatial tests.

On the other hand, domain-general accounts of working memory have been advanced by Engle and his colleagues (e.g. Engle, Tuholski, Laughlin & Conway, 1999). In such accounts, individual differences are interpreted in terms of the quantity of resources available. Turner and Engle's (1989) results led them to describe working memory as a general capacity resource, in which it is the capacity to keep active a certain number of elements that is crucial. However, and importantly from our perspective, they did not include a measure of spatial working memory in their study, so it is impossible to know whether their findings would generalise to the spatial domain. In any case, these two views are not incompatible. By "domain-general" Engle and colleagues mean that this capacity is not restricted to a certain type of task. Furthermore, in a later study, Kane, Hambrick, Tuholski, Wilhelm, Payne & Engle (2004) found that a two-factor model (in which verbal and visuo-spatial memory were separated) was a slightly better fit than a one-factor model in which working memory was regarded as a single construct, although the verbal and visuo-spatial working memory constructs were highly correlated. Thus, their data are consistent with a (weak) dissociation between verbal and visuo-spatial working memory capacity. In addition, Kane et al. provide some possible reasons to be sceptical of the data that purport to support strong domain specificity and we return to those reasons at the end of the introduction, since they are particularly pertinent to the design of our own study. In a meta-analysis of the relations between WM and comprehension, Carretti, Borella, Cornoldi and De Beni (2009) suggest both domain-general and specific factors play a role, with verbal working memory being more predictive. However, they compared verbal working memory only with visuo-spatial tasks (not numerical working memory) and only three of the studies they review included more than one type of working memory task.

Finally, there is still some ambiguity about the relation between numerical working memory tasks and comprehension in adults. For example, Waters and Caplan (1996) found that adults' comprehension was not significantly correlated with numerical working memory tasks, only with reading span tasks. In general, even if both sorts of task correlate with comprehension skill, it is the reading span tasks that show the stronger correlation.

In children, as opposed to adults, a number of studies have shown a strong relation between working memory and children's reading comprehension (e.g. Leather & Henry, 1994; Oakhill, Yuill & Parkin, 1986; Swanson & Berninger, 1995; Yuill, Oakhill & Parkin, 1989). This relation between working memory and reading comprehension has been found to hold with tasks that require the processing and storage of words (de Beni, Palladino, Pazzaglia & Cornoldi, 1998), sentences (Engle, Carullo & Collins, 1991; Seigneuric, Ehrlich, Oakhill & Yuill, 2000) and numbers (Yuill et al., 1989). Other studies have compared listening and counting span (Siegel & Ryan, 1989; Leather & Henry, 1994). Compared with the work on adults, however, there has been little research into domain-specificity of the relation between working memory and reading comprehension in children, and in particular the possible role of spatial working memory in children's comprehension.

Swanson (1992; 1996) argued for a general resources model, based on similar correlations between verbal and spatial working memory tasks and comprehension skill. However, this argument is not compelling and, indeed, other work by Swanson has produced less clear-cut results: Swanson and Berninger (1995) showed that, even with similar overall correlations between visuo-spatial working memory and comprehension skill, and verbal working memory and comprehension skill, verbal, but not visuo-spatial, working memory differentiated between groups of good and poor comprehenders. Thus, the issue of whether skilled reading comprehension in children is associated with general working memory remains equivocal, and will be taken up in the present study.

Bayliss and colleagues (Bayliss, Jarrold, Gunn & Baddeley, 2003; Bayliss, Jarrold, Baddeley, Gunn, & Leigh, 2005) also explored the relation between working memory and reading comprehension in children, using a sentence comprehension test (the NFER-Nelson group reading test II, 1998). They found moderate correlations between reading and both verbal and visuo-spatial span tasks (though not with a purely visuo-spatial task in their 2003 study). However, the reading comprehension measure was almost certainly confounded with word reading skills, which were not independently measured or controlled for. Indeed, in both studies, digit span was also correlated with the assessment of reading. This fact strongly suggests that the reading comprehension test was also assessing word reading which, unlike comprehension, tends to be associated with digit span.

An important, and novel, issue addressed in the present study is whether any of the working memory tasks are related to reading accuracy, as opposed to reading comprehension. We know of only two previous studies that explored the relation between reading comprehension and working memory in which assessments of reading comprehension skill were distinct from those of single word reading or decoding skills.

Seigneuric, et al. (2000) developed a test of spatial working memory: a simplified version of the tic-tac-toe task used by Daneman and Tardif (1987). They found that measures of working memory capacity – both verbal and numerical-predicted reading comprehension over and above vocabulary and decoding skills, but the spatial working memory task was not significantly related to comprehension skill. The present study builds on that of Seigneuric et al. in two important ways. First, we explore these relations over a wider age range and with more participants and, second, we control for general ability in each of the domains of interest.

The second study was conducted by Nation, Adams, Bowyer-Crain and Snowling (1999). The findings from their first experiment support those from previous studies (Oakhill et al., 1986; Stothard & Hulme, 1992) in showing that good and poor comprehenders do not differ in digit span and verbatim recall. Also in keeping with previous studies, Nation et al. found that good and poor comprehenders differed in verbal, but not in spatial, working memory and they argue that the poor comprehenders have specific problems with verbal processing and not more general capacity limitations. However, although Nation et al. collected data on the children's word reading accuracy, they did not look at the relation between working memory and word reading. The present study differs from theirs in various ways. First, we consider a wider age range, and include tasks of three different types (verbal, numerical and spatial). Second, Nation et al. compared the performance of groups who differed in comprehension ability on their working memory tasks, whereas we look at the relative contributions of working memory tasks in different domains, once general ability is controlled for. In the present study, we compared the relations of the various working memory tests to both accuracy and comprehension. We predicted that working memory would be more closely related to comprehension than to accuracy, and that it would predict variance in comprehension even when accuracy was controlled for.

A further issue addressed by our study is whether working memory systems become more differentiated with age. Contrasting views have been expressed by Alloway, Gathercole and Pickering (2006) and Hale, Bronik and Fry (1997), based, on the one hand, on correlational data and, on the other hand, on cross-task interference. It is possible that the relation between working memory and reading comprehension differs between children and adults. In particular, Kennedy and Murray (see, e.g. Kennedy, 1987; Murray & Kennedy,

1988) have suggested that spatial working memory is important for place-keeping skills in text comprehension, which might develop, at least partially, separately from other aspects of comprehension. These place-keeping skills of fluent readers allow them to re-inspect text selectively, and children who are good readers are much better at re-inspecting text selectively than poor readers (see Cataldo & Oakhill, 2000). Although spatial working memory does not predict text comprehension in adults, it might predict comprehension in children when these skills that depend on spatial working memory are developing, since the demands on the relevant memory systems may be higher. In this study, we assess the role of spatial and other measures of working memory in the reading comprehension performance of 6- to 8- and 9- to 11-year-olds separately.

In addition to the problems that we mentioned above, there is a more general problem of interpreting the relation between tests of working memory and assessments of reading. Working memory tests in any modality inevitably require some basic abilities in that domain. For example, reading and listening span tasks require general vocabulary knowledge. We might expect tests of verbal working memory to correlate better with measures of reading than tests of numerical working memory because of this shared dependence on general verbal ability. Of the previous studies, Yuill et al. (1989) used contrasting groups matched on basic vocabulary skills, but only Seigneuric et al. (2000) and Swanson (1992) have directly assessed this possibility by controlling for vocabulary or other general skills. Of course, tests of general ability are also likely to require some degree of working memory skills, but it would be an important indicator of the importance of working memory in reading if correlations with reading skills remained significant after performance on tests of general ability had been partialled out. In the present study, we include assessments of general ability in the three areas of interest: verbal, numerical and spatial, so that the particular working memory tasks can be assessed against the contribution of general ability in the relevant domain.

Another approach to this confounding was adopted by Yuill, Oakhill and Parkin (1989) who developed a working memory test that required processing and storage of numbers rather than words and sentences. They found a significant correlation between performance on this test and reading comprehension in 7- to 9-year-old children, but they did not directly compare the predictive power of their numerical task with that of a listening or reading span task, which is one of the aims of the present study.

A subsidiary question is how the level of verbal complexity of a working memory task contributes to the relation between that task and reading comprehension. Reading span tasks are complex, in that they require the simultaneous use of several skills: not only the verbal encoding of information and switching between storage and processing, but also the syntactic and semantic processing of sentences and the processing of word meanings (although the complexity of the processing component does not seem to relate very directly to the predictive power of the task: see Lépine, Barrouillet & Camos, 2005). In the present study, we developed and validated a verbal measure for use with children that does not require sentence-level comprehension, and compared its predictive power with that of the listening span task. In addition, the inclusion of these two verbal tasks enabled us to explore how tasks with different processing components relate to comprehension skill. We also included two different numerical tasks for similar reasons, but also because the final digit task we have used previously (see Yuill, et al., 1989), is not so strongly related as to comprehension skills as are verbal tasks (Seigneuric et al., 2000). One possible explanation for this weaker relation is that the processing requirement of the final digit task is low (children simply have to read out a set of single-digit numbers). In the present study we therefore also included a second test of numerical working memory, with a more demanding processing requirement. We used only one spatial task, a version of that used successfully in the study by Seigneuric et al.

Kane et al. (2004) point out some major limitations in previous (adult) studies that purport to show domain-specificity in working memory tasks, and we have attempted to overcome these criticisms in the present study, as follows. First, many studies have used small and quite homogeneous samples. We have used a large sample of children, across a wide age range. Second, in some studies, the verbal and spatial working memory tasks differ markedly in difficulty. We have piloted and developed tasks that were similar in difficulty, and adapted the level of difficulty to give similar levels of average performance in the different age groups. Third, it is not clear in previous studies whether it is the domain-specificity of the working memory construct that is important, or the domain-specificity of resources in that domain that are not specific to working memory tasks: for example, if verbal working memory is related to reading comprehension, is that something to do with the working memory task, or with the verbal nature of the task? In contrast to most previous studies of the dissociation between verbal and spatial working memory, we also included general measures of verbal, mathematical and spatial ability, so that the contributions of domain-specific working memory, as opposed to competence in a domain more generally, could be taken into account.

In summary, the present study aimed to explore the relation between working memory skills in different domains (verbal, numerical and spatial) and reading comprehension in children. This work extends previous studies in two main ways. First, we include comparisons of all three areas. Second, we explore the relation between working memory and reading skill when general ability in the relevant domain has been taken into account. Third, we explore whether any links between working memory and reading ability are specific to reading comprehension, or apply to reading skill more generally. Finally, we explore the way in which the level of complexity of the verbal and numerical tasks contributes to the relation between that task and reading skill.

Methods

Participants

All available children (excluding those few identified by teachers as having language or behavioural problems) from 12 classes of 6 to 11-year-olds in 5 schools took part in the study. This produced a sample of 197 children, divided into two age groups: 97 6- to 8- year-olds and 100 9- to 11-year-olds.

Overview of Design and Procedure

Each child completed 3 types of test, described in detail in the Materials section:

(1) reading tests (RA: Accuracy and RC: Comprehension tests of the Neale Analysis of Reading Ability Revised (Neale, 1997), (2) working memory tests of three types: verbal (2 tests, odd word out, VWM1, and reading span, VWM2), numerical (2 tests, highest number, NWM1, and final number, NWM2) and spatial (1 test, SPWM) and (3) general ability tests selected from the Cognitive Abilities Test (CAT: Thorndike & Hagen, 1986) in three areas: verbal, numerical and spatial (nonverbal).

The reading test was presented individually in the first session, followed by the 5 working memory tests, in a separate random order for each subject, spread over 3 sessions, each in a separate room by a male experimenter familiar to the children. Finally the CATs were presented to children in separate random orders in groups of 36.

Materials

Reading test. The Neale Analysis requires children to read aloud a series of narrative passages of increasing difficulty, and then to answer from memory a mix of factual and inferential questions about each passage. Children are corrected on words that are misread or not read, so that they are not disadvantaged on comprehension questions, but testing is stopped when children make a pre-set number of reading errors on a particular passage. Separate norm-referenced scores are computed for accuracy (number of words read correctly) and comprehension (number of questions correct). This test, and its predecessor, have been shown to predict a range of differences in abilities between good and poor comprehenders (e.g. see Yuill & Oakhill, 1991). We used the raw scores for reading accuracy (RA) and reading comprehension (RC) in all analyses.

Working memory tests. All five of these tests had certain characteristics in common. All required the simultaneous storage and processing of information. For each of the tests, there were four levels of storage difficulty presented in order of increasing difficulty, each level containing three trials (except the pre-existing final number task devised by Yuill et al., which was the least demanding in terms of storage and processing, and had eight trials). The first storage level contained two recall items and for each of the next levels the number of recall items was increased by one, with the final storage level having a maximum of five recall items. Where appropriate (in tests V1, V2, N1 and SP) the position of correct responses was counterbalanced. Children practised at each storage level until it was clear they understood the processing requirements of the task. This never required more than three trials at any of the storage levels.

In all of the tests a strict scoring procedure was used: children were required to recall the correct items in the order of presentation. In all analyses, we used the proportion of items correct out of the possible maximum score as the independent variable.

We piloted the materials for the new or adapted tests (V1, V2, N1 and SP) with 40 6- to 11-year-olds, in order to ensure that children could provide the correct responses for the processing component, and only used items on which 90% of the youngest children were correct. (Recall of these correct responses was, naturally, considerably less than 90%.) The second aim of the pilot was to ensure that the tests produced similar mean scores and standard deviations both across the different modalities and across the age range. This aim was achieved, both in the pilot and in the main study, as shown in Table 1.

In all the working memory tests, the older children were presented with more trials in order to ensure that the tasks were at an appropriate level of difficulty. The younger age groups received three trials at each of three levels (two, three or four items to recall) and the older children received an additional set of trials with five items to recall. Thus, the younger children were given a total of nine trials (3 trials at each of 3 levels of difficulty) and the older children received a total of 12 trials (3 trials at each of 4 levels of difficulty). All children attempted all trials appropriate for their age group. Because different children received different numbers of trials, the scores entered into the analyses were proportions correct. The exception was the spatial working memory task, which proved to be sufficiently difficult for all children with a recall demand of four.

Aural word span: Odd word out (VWM1). This newly-devised test consists of series of single words of one or two syllables in groups of four. Three of the words are in the same category (e.g. names of fruits or colours) and the fourth is from a different category. The words within each group of four are presented in a fixed random order. Children listen to the four words and have to detect the 'odd word out'. They then have to recall the odd words in each series.

An example of a three-item series is:

Whale	Shark	Dolphin	Scarf
Cowboy	Curtain	Indian	Sheriff
Egg	Aunt	Cousin	Uncle

The correct response for this series is: "scarf, curtain, egg".

Aural reading span (VWM2). This test is our UK English adaptation of the test used by Siegel and Ryan (1989). We adapted the test because we found that English children in some cases did not give the same completions as the original North American sample did. The child listens to a set of unrelated sentences and supplies the final word in each, then recalls these final words. The final words are highly constrained by the context. For example,

The sun shines during the day, the moon at _____.

At the library people read ______.

An apple is red, a banana is ______.

The correct response is: "night, books, yellow".

Highest Number Task (NWM1). In this new test, children inspect sets of three numbers, shown on a card and read aloud by the experimenter. They have to pick the highest number and then recall the highest numbers from each set. All the numbers are between 1 and 19 and each set contains one number below 10, and two between 10 and 20. For example, a 3-item set is:

14	9	17
10	11	4
15	3	12

to which the answer is: "17, 11, 15".

Final Number Test (NWM2). This task was the one developed by Yuill et al. (1989). Children are required to read sets of three-digit numbers and to recall the last numeral in each number. For example, a three-item set is:

528

434

489

to which the answer is: "8, 4, 9".

It should be noted that, though we refer to the two above tests as tests of "numerical" working memory, they might better be described as tasks that require numerical processing, but verbal storage. We return to this point in the Discussion section.

Spatial Working Memory Test (SPWM). Daneman and Tardif (1987) described a spatial test using three-dimensional tic-tac-toe. We adapted this test to make it suitable for children. Children were shown a series of 3 by 3 matrices, one at a time, each containing two noughts, and had to point to the cell where a nought should be inserted to make a winning line. Each grid has noughts of a different colour, in order to facilitate children's recall of the positions in the correct order. After seeing all the matrices, children had to place strips of corresponding colours onto an adhesive grid, to indicate the positions of the winning lines. The colours of the lines could thus be linked to the order in which the grids had been presented: the sequence of colours was the same in each trial (e.g. the trials of 3 items showed orange, then blue, then green noughts, and the trials of 4 items added a set of pink noughts to this sequence). Although the probability of guessing the correct lines is 1 in 8 (as there are only 8 possible winning lines in a 3 by 3 grid), the probability of matching colours to positions is much lower.

Cognitive Abilities Tests (CATs). These are a series of standardised pencil-and-paper multiple-choice tests tapping general ability in different areas. The full CAT consists of four verbal subtests, three mathematical ('quantitative') subtests and three spatial ('nonverbal') subtests. We used one of each type of subtest: Verbal 2, Quantitative 1 and Nonverbal 2, all preceded by two to three practice items. These subtests were chosen on the basis that each correlated most highly with the other subtests in its battery and was most highly loaded on the relevant factor in a factor analysis (Thorndike, Hagen & France, 1986). There are two levels of each test: Level A consists of the first 25 questions in the test, which increase in difficulty through the test, and level B consists of the first 30 questions of the same test. All children took level B, except for the 7-year-old group, who were given Level A. An example from each of these tests is given below.

Verbal (CATV):

The fire is .... (possible choices: wet, green, hot, running, round)

Quantitative (CATN):

Which is greater: 1 1 + 1 1 OR 1 1 + 1 ?

Spatial (CATS):

Respondents have to make analogies between diagrams, for example: Large square is to small square as large circle is to ....?, with choices of a small circle, a large semicircle, a filled square and two triangles joined at the apices.

Results

Descriptive Statistics

Each child had a maximum of 10 test scores (though because of time constraints and absences, it was not always possible to collect the full set of data for every child; where data were missing, correlations and regression analyses were conducted using all available data: the actual numbers of children included in the correlation analyses is shown in the relevant tables): two reading scores, Neale Reading Accuracy (NRA) and Neale Reading Comprehension (NRC), five working memory scores, (Verbal, Odd Word; Verbal Reading Span; Numerical Highest Number; Numerical Final Digit and Spatial) which were calculated as proportion of total possible score, and three general ability scores (CAT Verbal, CAT Numerical and CAT Spatial). We also calculated chronological age at time of test in months (CA). The means and standard deviations of each score for the two age groups are shown in Table 1. The working memory tests, including the new ones, showed a reasonable spread of performance (all means within the range 35-59% correct) and similar levels of variability (alls s.d.s in the range .12 to .16). Importantly, the tests that turned out to be the strongest predictors of comprehension skill were not differentiated from the other tests by their particular ease or difficulty.

Table 1. *Summary of Means and Standard deviations of Overall Scores on the Assessments, and within each Age-group.*

	Age	Neale Accuracy	Neale compreh.	CAT: verbal	CAT: numerical	CAT: spatial	VWM: odd word	VWM: reading span	NWM: highest number	NWM: final digit	Spatial WM
Age group:											
6-8years Mean	94.84	36.70	12.38	12.07	11.18	12.73	.35	.52	.44	.48	.44
N	83	84	84	82	82	81	78	81	77	80	75
s.d.	6.98	17.64	5.32	5.72	5.71	8.29	.13	.14	.14	.14	.14
9-11 years Mean	121.10	62.36	20.95	21.20	21.05	20.84	.46	.59	.53	.55	.53
N	95	95	95	95	94	95	66	81	65	77	83
s.d.	7.53	21.59	7.73	5.68	5.44	8.92	.12	.12	.16	.14	.15
Overall Mean	108.85	50.32	16.93	16.97	16.45	17.11	.40	.55	.48	.51	.49
N	178	179	179	177	176	176	144	162	142	157	158
s.d.	14.73	23.45	8.19	7.03	7.36	8.54	.12	.14	.16	.14	.15

Note: Chronological age is in months; Neale and CAT scores are raw scores; Working Memory Tasks are proportion correct.

The reliability of the different working memory measures was calculated using Cronbach's Alpha. It will be recalled that older children were required to complete more trials to ensure that the tasks were sufficiently difficult for them., so the total number of trials was 12 for the older children and 9 for the younger ones. However, estimates of reliability could be obtained only over the items that were completed by all the participants (*n* = 9 items) and, given the small number of items, were acceptable. The levels of Cronbach's Alpha ranged between .66 and .73 for the verbal and numerical tasks, but were slightly lower for the spatial working memory task (.61). It was not appropriate or necessary to calculate reliability in the case of the Neale Analysis scores or the CATs scores, since these are standardised tests with published reliability statistics.

Correlational analyses

The bi-variate correlations between the measures are shown overall, and separately for each age group, in Table 2. Because of the large number of correlations, we adopted a conservative (.01) level of significance. All five working memory measures were significantly correlated with both reading comprehension (correlations between .34 and .46) and reading accuracy (correlations between .36 and .47), and all the working memory measures were significantly correlated with each other (correlations between .34 and .59). Some of these correlations held up in both age groups separately, but it was only the Verbal Reading Span and Numerical Final Digit Task that were strong and consistently related to reading comprehension in both age groups. We return to this point in the regression analyses, presented below. In addition, the general ability measures (CATs) were significantly correlated with both the working memory measures and the reading measures in the data overall, and the CATs scores were correlated with the two reading ability measures in both age groups, with the exception of the CAT Spatial which did not correlate with RC in the younger group. The relation between the CATs scores and the working memory assessments in the two age groups considered separately were less consistent.

Table 2. *Intercorrelations among Age and Ability Variables*

		Age	Neale Accuracy	Neale Comprehension	VWM: odd word	VWM: reading span	NWM: highest number	NWM: final digit	Spatial WM:	CAT: verbal	CAT: Numerical	CAT: Spatial
Age		1	.602**	.581**	.427**	.250**	.334**	.211**	.413**	.716**	.734**	.483**
	N		178	178	143	162	141	157	158	176	175	175
Neale RA			1	.713**	.470**	.374**	.421**	.400**	.359**	.813**	.711**	.578**
	N			179	144	162	142	157	158	177	176	176
Neale RC				1	.464**	.464**	.345**	.440**	.344**	.715**	.612**	.520**
	N				144	162	142	157	158	177	176	176
VWM1					1	.590**	.515**	.506**	.470**	.492**	.464**	.488**
	N					133	139	131	129	142	141	141
VWM2						1	.472**	.542**	.339**	.377**	.369**	.344**
	N						131	157	146	160	159	159
NWM1							1	.559**	.483**	.394**	.511**	.389**
	N							127	126	140	140	139
NWM2								1	.359**	.361**	.463**	.344**
	N								141	155	154	154
SPWM									1	.384**	.454**	.439**
	N									156	155	155
CAT-V										1	.772**	.666**
	N										176	176
CAT-N											1	.592**
	N											175

** p< 0.001

* p < 0.01

Table 2a. *Intercorrelations among Variables within the Younger Age Group*

		Age	Neale Accuracy	Neale Compreh.	VWM: odd word	VWM: reading span	NWM: highest number	NWM: final digit	Spatial WM	CAT: verbal	CAT: numerical	CAT: spatial
Age		1	.300*	.215	.271	.267	.210	.056	.328*	.580**	.408**	.216
	N		83	83	77	81	76	80	75	81	81	80
Neale RA			1	.652**	.385**	.458**	.228	.456**	.273	.739**	.523**	.299*
	N			84	78	81	77	80	75	82	82	81
Neale RC				1	.428**	.466**	.280	.406**	.184	.580**	.348*	.244
	N				78	81	77	80	75	82	82	81
VWM 1					1	.613**	.503**	.451**	.500**	.433**	.315*	.491**
	N					75	77	74	71	76	76	75
VWM 2						1	.501**	.427**	.391**	.403**	.344*	.308*
	N						74	80	74	79	79	78
NWM 1							1	.372**	.398**	.241	.369**	.290
	N							73	70	75	75	74
NWM 2							1	.372**	.398**	.241	.369**	.290
	N							73	70	75	75	74
SPWM									1	.249	.291	.285
	N									73	73	72
CAT-V										1	.585**	.466**
	N										82	81
CAT-N											1	.355**
	N											81

** p< 0.001
* p < 0.01

Table 2b. *Intercorrelations among Variables within the Older Age Group*

		Age	Neale Accuracy	Neale Compreh.	VWM: odd word	VWM: reading span	NWM: highest number	NWM: final digit	Spatial WM	CAT: verbal	CAT: numerical	CAT: spatial
Age		1	.314*	.312*	-.030	-.176	.119	-.065	.337*	.339**	.441**	.268*
	N		95	95	66	81	65	77	83	95	94	95
Neale RA			1	.567**	.295	.132	.410**	.224	.233	.715**	.586**	.560**
	N			95	66	81	65	77	83	95	94	95
Neale RC				1	.242	.369**	.200	.368**	.262	.587**	.453**	.464**
	N				66	81	65	77	83	95	94	95
VWM 1					1	.446**	.399**	.413**	.304	.216	.261	.284
	N					58	62	57	58	66	65	66
VWM 2						1	.353*	.615**	.181	.159	.200	.235
	N						57	77	72	81	80	81
NWM 1							1	.688**	.481**	.331*	.505**	.329*
	N							54	56	65	65	65
NWM 2								1	.332*	.273	.427**	.212
	N								68	77	76	77
SPWM									1	.289*	.430**	.419**
	N									83	82	83
CAT-V										1	.637**	.647**
	N										94	95
CAT-N											1	.537**
	N											95

** p< 0.001

* p < 0.01

Since our prediction in relation to age differences was not upheld (that there may be a different relation between comprehension skill and visuo-spatial working memory in the two age groups) we conducted all further analyses on the entire data set.

Regression analyses

The regression analyses enabled us to assess the relative importance and specificity of the various predictors in relation to the measures of reading ability. We were particularly interested in comparing the predictive power of the numerical, verbal and spatial tasks.

The first goal was to determine whether working memory was a predictor of reading comprehension when age and general ability in the relevant domain were controlled. A first set of stepwise hierarchical multiple regression analyses were conducted with comprehension as the independent variable, in which the different working memory measures were entered at the final step. In each analysis, three variables were entered: age, performance on the relevant Cognitive Abilities Test (Verbal, Numerical or Spatial, depending on which working memory task was entered) and one of the working memory tasks. Thus, five different models were tested – each with a different working memory measure.

In all of these analyses, age and the relevant CAT were highly significant predictors of comprehension skill. However, the results showed that three of the four verbal and numerical working memory tests (but not the spatial test) accounted for variance in comprehension skill over and above that accounted for by age and the relevant CAT score. The Reading span task and the Final Digit task were the strongest predictors of comprehension skill, and the other verbal task (Odd Word Out) was only marginally predictive. The results of these regression analyses are shown in Table 3.

Table 3: *Stepwise Multiple regression Analyses Predicting Reading Comprehension.*

Table 3a: *VWM1 (odd word) Entered in Final Position*

Independent Variable	*R Square*	*R Square Change*	*F Change*	*d.f.*	*Sig. F Change*
1. Age	.306	.306	61.222	1,139	.001
2. CAT-V	.534	.228	67.589	1,138	.001
3. VWM 1	.546	.012	3.684	1,137	.057

Table 3b: VWM2 (reading span) Entered in Final Position

Independent Variable	*R Square*	*R Square Change*	*F Change*	*d.f.*	*Sig. F Change*
1. Age	.327	.327	76.710	1,158	.001
2. CAT-V	.506	.180	57.128	1,157	.001
3. VWM 2	.553	.047	16.371	1,156	.001

Table 3c: *NWM1 (highest number) Entered in Final Position*

Independent Variable	*R Square*	*R Square Change*	*F Change*	*d.f.*	*Sig. F Change*
1. Age	.316	.316	63.342	1,137	.001
2. CAT-N	.397	.081	18.277	1,136	.001
3. NWM 1	.399	.002	.486	1,135	.487

Table 3d: *NWM2 (final digit) Entered in Final Position*

Independent Variable	*R Square*	*R Square Change*	*F Change*	*d.f.*	*Sig. F Change*
1. Age	.320	.320	71.644	1,152	.001
2. CAT-N	.389	.068	16.854	1,151	.001
3. NWM 2	.444	.056	15.007	1,150	.001

Table 3e: *Spatial Working Memory Span Entered in Final Position*

Independent Variable	*R Square*	*R Square Change*	*F Change*	*d.f.*	*Sig. F Change*
1. Age	.332	.332	76.172	1,153	.001
2. CAT-Sp	.421	.089	23.375	1,152	.001
3. SpWM	.422	.001	.250	1,151	.618

Thus far, the results closely parallel those of Seigneuric et al. but also go beyond them in important ways, in that we control for measures of general ability in the relevant domain, whereas they did not. That is, even after performance on the relevant assessment of general ability measure had been entered, both verbal working memory measures, and the numerical final digit measure accounted for significant variance in comprehension skill. It is particularly impressive that the tests of verbal working memory accounted for variance in comprehension skill over and above the contribution of general verbal ability, since that variable alone accounted for around 20% of unique variance in comprehension skill.

Since the reading scores were highly correlated in this sample (r = .71) and each of the working memory tasks was correlated with accuracy in the sample overall (all $rs \geq .36$) it is important to establish whether working memory is a predictor of comprehension skill specifically, or reading more generally. In order to do this, we conducted a parallel set of regression analyses to those above, with reading accuracy as the dependent variable. After controlling for chronological age and the relevant measure of general ability, only one of the working memory measures predicted significant variance in word reading. That was the final digit task, which accounted for 1.6% of variance in accuracy (compared with 5.6% in comprehension). Because at least one of the working memory tasks was related to reading accuracy, over and above the effects of age and the general ability measure, we re-ran the regression analyses in which comprehension was the dependent variable, but controlled for reading accuracy as well as chronological age and the relevant general ability measure. Despite this very strong test of the predictive power of the measures of working memory, the verbal (Reading span) and numerical (Final Digit) working memory measures continued to predict variance in comprehension skill.

This first set of analyses enables us to provide a clear answer to the first question, which is whether working memory predicts comprehension skill in children over and above measures of general ability. The answer is that three of the four verbal and numerical working memory measures account for significant (or marginally significant) variance in comprehension skill, over and above the effects of age and a relevant general ability measure. It replicates and extends Seigneuric et al.'s (2000) finding that spatial working memory was not related to comprehension skill.

The second question concerns the nature of the working memory resources involved in reading comprehension. We wanted to determine whether the working memory system that is related to reading comprehension in children is a general system, or a symbolic system specialised for language processes. The results of the previous analyses provide some indications. As we saw above, only the verbal working memory tasks and one of the numerical tasks were significantly related to comprehension skill in both age groups once age and general ability had been partialled out, whereas the spatial working memory task was not related to comprehension skill over and above age and general ability. Thus, these analyses seem to support the "symbolic resource model" that we describe in the Introduction. In order to test this hypothesis more directly, we need to assess whether the verbal and numerical tasks draw on the same pool of symbolic resources to predict reading comprehension. Therefore we carried out a further set of analyses.

In these analyses, performance on the stronger verbal working memory task (Reading span) and the stronger numerical working memory task (Final Digit) was compared. The variables were again entered in a fixed order, and the order of entry of the verbal and the numerical working memory measures was reversed in order to assess the shared and the unique variance explained by each measure. Support for the symbolic system hypothesis would come from results showing the contribution of a verbal or numerical task to be substantially reduced when the effect of the other (numerical or verbal) task was previously entered into the regression equation. In these analyses, the spatial task was not considered further since it did not account for significant variance in reading comprehension, over and above age and a general measure of spatial ability. Because of the wide age range, and the general improvement with age on the working memory tasks, age was entered first in the regression analyses. It was not obvious which of the general ability measures (verbal or numerical) to enter in these analyses, but in fact the results showed an identical pattern whichever was used. The results including the Verbal CAT data are presented.

The results of these regression analyses are shown in Table 4. These analyses indicate that the verbal and numerical measures contributed independently to variance in comprehension skill, even after controlling for domain-relevant ability. We note here that this result is contrary to that obtained by Seigneuric et al., who found that neither of their numerical tasks explained variance over and above that contributed by one of the verbal measures. However, our data are probably more reliable since the sample was much larger (in these particular analyses 153 participants were entered as opposed to only 48 in Seigneuric et al.'s study). Despite the apparent difference in conclusion, however, we found that whichever working memory assessment was entered last contributed a very small percentage of additional variance (between 1.5 and 1.7%, see Table 4) though, of course, the preceding variables had already taken up about 55% of the variance. Thus, although the contribution of whichever task is entered at the final step is significant, there is also a very substantial amount of shared variance between the verbal and numerical working memory tasks.

Table 4: *Fixed Order Regressions with the Verbal (Reading Span) and a Numerical (Final digit) Working Memory Measure as Predictors; Reading Comprehension as the Dependent Variable*

Independent Variable	*R Square*	*R Square Change*	*F Change*	*d.f.*	*Sig. F Change*
1. Age	.321	.321	72.355	1,153	.001
2. CAT-V	.505	.184	56.505	1,152	.001
3. VWM 2	.550	.045	15.148	1,151	.001
4. NWM 2	.565	.015	5.109	1,150	.025
3. NWM 2	.548	.043	14.309	1,151	.001
4. VWM 2	.565	.017	5.896	1,150	.016

A subsidiary question, which we addressed in a further analysis, was the way in which the level of verbal complexity of a working memory task contributes to the relation between that task and reading comprehension. Reading span requires not only switching between storage and processing, verbal encoding of information and phonological storage, but also requires the syntactic and semantic processing of sentences and the processing of word meanings. The odd-word-out task was designed to tap the same processes, except that it did not require sentence processing. Thus, if it is the shared verbal component of the tasks that is important in predicting comprehension, then we might expect that the reading span task would not predict additional variance over and above that predicted by the odd-word-out task. However, if the sentence-level processing is an important additional aspect of the predictive power of the reading span task, over and above the more general verbal component, then we might expect that it would account for additional variance even after performance on the odd-word-out task is controlled for. In the analyses in which we compared the predictive power of the two verbal tasks, we found the latter pattern of results (see Table 5). As can be seen, the reading span task accounted for an additional, highly significant, 3% of variance even when entered last, whereas the word span task, when entered last, did not account for significant additional variance

Table 5: *Fixed Order Regression Analyses to Compare the Predictive Power of the Two Verbal Working Memory Tests (Reading Span and Odd Word Out)*

Independent Variable	*R Square*	*R Square Change*	*F Change*	*d.f.*	*Sig. F Change*
1. Age	.312	.312	58.592	1,129	.001
2. CAT-V	.535	.223	61.372	1,128	.001
3. VWM 1	.552	.017	4.740	1,127	.031
4. VWM 2	.580	.028	8.420	1,126	.004
3. VWM 2	.579	.044	14.309	1,127	.001
4. VWM 1	.580	.001	5.896	1,126	.608

Overall, these results provide support for the idea that working memory capacity is a strong predictor of reading comprehension in children. Both verbal and numerical (but not spatial) working memory tasks contributed substantially to the prediction of reading comprehension even when age and relevant general ability had been taken into account, but there was some indication that they are making independent contributions to this prediction. The results also provide some support for the idea of a symbolic capacity model since the verbal and numerical tasks were far more strongly related to comprehension skill than was performance on the spatial task.

Discussion

For the purposes of this study, we produced working memory tests for children across a wide age range. The tests were calibrated in such a way that they produced similar means and standard deviations from 6 through to 11-year olds. The new tests we have developed (the odd-word out task, the largest number task and the spatial task) and the comparison of these with tests of working memory we have used previously (the reading span and final-number tests) provide useful comparative data, and also provide new evidence on the relation of working memory to children's reading skills: both comprehension and reading accuracy.

An important theoretical aspect of this study is that we have explored the predictive power of the various measures of working memory (verbal, numerical and spatial) once performance on the relevant general ability has been controlled. We found that, even after discounting age and relevant general ability, two of the tests (the reading span and final digit tasks) strongly predicted, and one (the odd-word-out task) marginally predicted, performance on the comprehension assessment. Of course, although we refer to two of the tests as "numerical tests", they do require verbal storage (it is only the processing component that is numerical), so it is not surprising they are related to verbal tests. Indeed, the numerical tasks were significantly correlated with the verbal tasks, and more highly than they were with the spatial tasks. The spatial working memory test was moderately correlated with both accuracy and comprehension overall, but not predictive of comprehension once age and spatial ability had been partialled out.

These findings are consistent with those of Leather and Henry, who found a reading span task to be a significant predictor of comprehension in 7-year-olds, while Yuill et al. (1989)

found a high correlation between comprehension skill in 7- to 8-year-olds and the same digit working memory test that we used in the present study. Our results suggest that, overall, the verbal tasks are better predictors than numerical ones, but that the difference is not striking. Indeed, there was little difference in the predictive power of the two stronger numerical and verbal working memory tests: both accounted for about 5% of unique variance in comprehension skill, over and above the contributions of age and the relevant measure of ability in that domain. This pattern of findings suggests that verbal tasks, even ones that require sentence comprehension, are not particularly privileged in their relation to reading comprehension. Importantly, it argues against the idea that reading span measures relate to comprehension skill only because they have a comprehension component (see also, Lépine, et al, 2005).

These conclusions are consistent with the suggestion from previous (adult) studies (e.g. Daneman & Tardif, 1987) that the working memory system that is implicated in language comprehension is a system specialized for the processing of symbolic (particularly verbal) information, and that spatial working memory does not have a role in text processing.

In the present study, we were able to compare the relative predictive power of two verbal working memory tests, which we designed to have different characteristics, crucially their level of verbal complexity. The reading span task requires not only switching between storage and processing, the verbal encoding of information and phonological short-term storage, but also the syntactic and semantic processing of sentences, including word meanings. The odd-word-out task, in contrast, does not require any sentence processing (though it does require knowledge of word meanings and categories since the child has to select and remember the word in each set that comes from a different semantic category). Thus, if the sentence-level processing is an important additional aspect of the predictive power of the reading span task, then we might expect that it would account for additional variance even after performance on the other verbal task (the odd-word-out task) was controlled for, and this is what we found. However, as we have argued above, any explanation of the role of WM in comprehension that focuses entirely on the comprehension requirement of the reading span task does not hold up, given that the final digit task was as strong a predictor of reading comprehension. Furthermore, a comparison of the two numerical tests showed that the final digit test was a considerably stronger predictor than the highest number task, even though the final digit task had a low processing requirement (reading out digits) and did not make demands on sentence comprehension processes. Thus, our more general prediction about the relation between working memory and reading comprehension being dependent on the level of processing difficulty was not supported by the data (see also Lépine et al., 2005).

The ability of one of the verbal and one of the numerical working memory tests to predict comprehension skill remained strong after tests of general ability had been partialed out. This method of analysis represents a very conservative test of the hypothesised relation between working memory and comprehension, since ability in a particular modality will inevitably influence children's performance on working memory tests in that modality. As we noted previously, tests of general ability, including the CAT, can in turn be expected to tap working memory capacity to a limited degree.

Importantly, the pattern of relations reported above was specific to comprehension and did not apply to reading accuracy. Although working memory in each of the different tasks was correlated with reading accuracy, both overall and in each age group separately, most of the predictive power was lost when age and the relevant ability score were first entered into the regression equation. Only one of the working memory tasks (the final digit task) predicted a very small proportion of variance in word reading accuracy over and above age and the relevant CAT. Our finding that performance on particular working memory tasks was related specifically to reading comprehension, and not word reading accuracy, is consistent with some previous results (e.g. Swanson & Jerman, 2007; Oakhill, Cain & Bryant, 2003; Yuill et al., 1989). Swanson and colleagues have demonstrated that reading comprehension and growth in reading comprehension are best predicted by executive function tasks (i.e. working memory tasks), rather than short-term memory tasks involving phonological coding, which are more likely to be related to word recognition skills.

Although spatial working memory and reading comprehension were correlated, spatial working memory was not predictive of reading comprehension once age and spatial ability had been partialled out. Consistent with the results of Nation et al., (1999) and Seigneuric et al. (2000), there is no evidence, even within this wide age group, that the spatial working memory system plays a role in comprehension processes, and neither is it related more strongly to comprehension skill in younger than in older children, as we hypothesized it might be. Although we used the same visuo-spatial task as Seigneuric et al., this was different to the task used by Nation et al., thus demonstrating that the pattern of results in children generalizes over more than one task. We have suggested that the different predictive power of the spatial vs. the numerical and verbal working memory tests might be explained in terms of the verbal and numerical tests requiring processing of symbolic information (letters and numbers). However, an alternative explanation of the difference, which might be considered in further work, is that the verbal and numerical tasks, unlike the spatial task, depend on retrieval of information from long-term memory (which is, of course, also a characteristic of skilled reading comprehension. We also suggested, following Kennedy and Murray, that spatial working memory skills might be specifically related to the skill of place keeping in comprehension. However, it could be that the comprehension assessment is not sufficiently demanding of these skills. In any case, the time course of typical uses of place holding skills in reading is perhaps beyond the bounds of what is traditionally thought of as working memory processes. Thus, such skills might be more closely linked to spatial ability, rather than working memory *per se,* which is consistent with our finding that general spatial ability was a better predictor of reading comprehension in our sample than was the measure of spatial working memory.

As outlined in the Introduction, working memory skills are likely to be important in reading (and listening) comprehension (in both children and adults). The processes of integration and inference are important to the construction of an integrated and coherent model of a text (i.e. a *Mental Model* or a *Situation Model:* Gernsbacher, 1990; Johnson-Laird, 1983; Kintsch, 1998), and these processes require that the relevant information, either from the text or world knowledge, is both available and accessible. Working memory is proposed to serve as a buffer for the most recently read propositions in a text, enabling their integration to establish coherence, and for information retrieved from long-term memory to enable its integration with the currently active text (see e.g. Cooke, et al., 1998; Graesser, et al., 1994).

This study, along with several previous studies, demonstrates a strong relation between memory and children's reading comprehension. The majority of this work suggests that the relation between memory and reading comprehension is specific to working memory tasks that require the simultaneous storage and processing of symbolic information (both verbal and numerical), rather than memory tasks that simply assess the passive storage of such information (e.g. Leather & Henry, 1994; Oakhill et al., 1986; Swanson & Berninger, 1995; Yuill et al., 1989). Furthermore, and consistent with the present findings, the working memory resources that are related to reading comprehension appear to be specialised for language processing: tasks that require the manipulation of shapes and patterns do not explain variance in reading comprehension skill (Nation et al., 1999; Seigneuric et al., 2000).

This work does not establish directly whether it is the controlled attention aspect of working memory, or the storage function (or STM) that is important in reading comprehension. Both STM and working memory deficits have been shown to make contributions to reading problems (e.g. de Jong, 1998; Swanson & Ashbaker, 2000). However, STM tasks do not discriminate between good and poor comprehenders who are matched for word recognition (e.g. Yuill & Oakhill, 1991), and Swanson & Jerman (2007) found that it was working memory and not STM that predicted growth in reading comprehension. Of course, working memory tasks have an STM requirement, but there must be something over and above this storage component that is important to comprehension skill and its development.

In addition, although working memory capacity assessed by symbolic processing tasks explains individual differences in children's text comprehension over and above other well established predictors of reading comprehension, such as word recognition skill and vocabulary knowledge (e.g., Swanson & Berninger, 1995; Yuill et al., 1989), some researchers have suggested that the reported relation between children's working memory and text comprehension is the result of underlying levels of verbal and semantic skills. For example, Nation et al. (1999) argue that poor comprehenders have a specific semantic weakness that restricts their ability to store verbal information in short-term memory and that this weakness, in turn, impairs their performance on verbally mediated working memory tasks. A similar position was adopted by Stothard and Hulme (1992), who proposed that working memory differences between good and poor comprehenders would disappear if differences in verbal IQ were controlled. The present results argue against this position, since at least two of the working memory tasks were strongly predictive of comprehension skill over and above the contribution of general verbal (or numerical) ability, and demonstrate that working memory tasks that require symbolic processing are important predictors of comprehension skill in young children, over and above the cognitive skills that contribute to the tasks.

• • •

Jane Oakhill, is Professor of Experimental Psychology at the University of Sussex, Brighton, UK. She worked briefly as a primary school teacher, and her interest in children's reading comprehension problems led to a D.Phil. on that topic. She has worked on various research projects (including deductive reasoning in children and adults, circadian variations in human performance, and adult language comprehension) but has always maintained a research interest in children's reading comprehension (in particular, individual differences).

Nicola Yuill is head of the Children and Technology Lab (ChaTLab) at the University of Sussex, UK. She has worked for many years on children's reading comprehension, especially intervention, and studies the role of social interaction and collaboration, particularly supported by technology, in understanding and developing language and communication in typical and atypical populations.

Alan Garnham is Professor of Experimental Psychology in the School of Psychology at the University of Sussex, UK, where he has worked for most of his professional career. His research interests are primarily in text comprehension, in particular anaphor processing. He has also worked on syntactic processing and on deductive reasoning.

References

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*, 1698-1716.

Baddeley, A.D. & Hitch, G.J. (1974). Working memory. In G.A. Bower (Ed.) *The psychology of learning and motivation,* Vol. 8 (pp. 4790). New York: Academic Press.

Bayliss, D. M., Jarrold, C., Baddeley, A. D., Gunn, D. M., & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Developmental Psychology, 41,* 579-597.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132*, 71-92.

Caplan, D. and Waters, G. S. (1999) Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences, 22,* 77-126

Cantor, J., Engle, R.W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence, 15,* 229-246.

Carretti, B., Borella, E., Cornoldi, C. & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences, 19,* 246–251

Cataldo, M.G. and Oakhill, J.V. (2000) Why are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial memory on ability to locate information in a text. *Journal of Educational Psychology, 92*, 791-799.

Cooke, A. E., Halleran, J. G., & O'Brien, E. J. (1998). What is readily available during reading? A memory-based view. *Discourse Processes, 26,* 109-129.

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

Daneman, M. & Carpenter, P.A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9,* 561-584.

Daneman, M. & Merikle, P.M. (1996). Working memory and language comprehension: A metaanalysis. *Psychonomic Bulletin & Review, 3,* 422-433.

Daneman, M. & Tardif, T. (1987). Working memory and reading skill reexamined. In M. Coltheart (ed.), *Attention and performance,* 7.Pp. 491-508. London: Erlbaum.

de Beni, R., Palladino, P., Pazzaglia, F., & Cornoldi, C. (1998). Increases in intrusion errors and working memory deficit of poor comprehenders. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 51A,* 305–320.

de Jong, P. (1998). Working memory deficits of reading disabled children. *Journal of Experimental Child Psychology, 70,* 75–95.

Dixon, P., LeFevre, J. & Twilley, L.C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology, 80,* 465-472.

Engle, R. W., Carullo, J. J., & Collins K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research, 84,* 253-262.

Engle, R. W., Nations, J. K., & Cantor, J. (1990). Is "working memory capacity" just another name for word knowledge? *Journal of Educational Psychology, 82,* 799-804.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999b). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 125,* 309 – 331.

Gernsbacher, M. A. (1990). *Language Comprehension as Structure Building*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Hale, S., Bronik, M. D., & Fry, A. E. (1997). Verbal and spatial working memory in school-age children: Developmental differences in susceptibility to interference. *Developmental Psychology, 33,* 364–371.

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

Just, M. A., & Carpenter P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122-149.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189-217.

Kennedy, A. (1987). Eye movements, reading skill and the spatial code. In J.R. Beech and A.M. Colley (Eds.) *Cognitive Approaches to Reading.* Chichester: Wiley.

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

Leather, C.V. & Henry, L.A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology, 58,* 88-111.

Lépine, R., Barrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychonomic Bulletin and Review, 12,* 165-170.

Murray, W.S. & Kennedy, A. (1988). Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye movement analyses. *Quarterly Journal of Experimental Psychology, 1988, 40A,* 693-718.

Nation, K., Adams, J. W., Bowyer-Crane, C.A. & Snowling, M. J. (1999) Working Memory Deficits in Poor Comprehenders Reflect Underlying Language Impairments. *Journal of Experimental Child Psychology, 73,* 139-158.

Neale, M.D. (1997) *The Neale Analysis of Reading Ability: Revised,* London: Macmillan Education.

Oakhill, J. V., Cain, K.E. & Bryant, P.E. (2003). Dissociation of single-word reading and text comprehension skills. *Language and Cognitive Processes, 18,* 443-468.

Oakhill, J. V., Yuill, N. M., & Parkin, A. (1986). On the nature of the difference between skilled and less-skilled comprehenders. *Journal of Research in Reading, 9,* 80–91.

Seigneuric, A., Ehrlich, M-F., Oakhill, J.V. & Yuill, N.M. (2000). Working memory resources and children's reading comprehension. *Reading and Writing, 13,* 81-103.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125,* 4-27.

Siegel, L.S., & Ryan, E.B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development, 60,* 973-980.

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.

Swanson, H.L. (1992) Generality and modifiability of working memory among skilled and less skilled readers. *Journal of Educational Psychology, 84,* 473-488.

Swanson, H.L. (1996). Individual and age-related differences in children's working memory. *Memory & Cognition, 24,* 70-82.

Swanson, H. L., & Ashbaker, M. (2000). Working memory, short-term memory, articulation speed, word recognition, and reading comprehension in learning disabled readers: Executive and/or articulatory system? *Intelligence, 28,* 1–30.

Swanson, H.L., & Berninger V.W. (1995). The role of working memory in skilled and less skilled readers' comprehension. *Intelligence, 21,* 83-108.

Swanson, H.L. & Jerman, O. (2007) The influence of working memory on reading growth in subgroups of children with reading disabilities. *Journal of Experimental Child Psychology , 96,* 249–283

Thorndike, R.L. & Hagen, E. (1986) *Cognitive Abilities Test Levels AF*. Second Edition. NFER Nelson: Windsor, Berks.

Thorndike, R.L., Hagen, E. & France, N. (1986) *Cognitive Abilities Test: Administration Manual.* Second Edition. NFER-Nelson: Windsor, Berks.

Turner, M.L., & Engle, R.W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127-154.

Waters, G.S. & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology, 1996, 49A,* 51-79.

Yuill, N. & Oakhill, J. (1991). *Children's Problems in Text Comprehension: An Experimental Investigation.* Cambridge: Cambridge University Press.

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

International Electronic Journal of Elementary Education, 2011, 4(1), 107-125.

The effects of syntactic and lexical complexity on the comprehension of elementary science texts

Diana J. ARYA*
University of Oslo, Norway

Elfrieda H. HIEBERT
TextProject and University of California, Santa Cruz, United States

P. David PEARSON
University of California, Berkeley, United States

Abstract

In this study we examined the effects of syntactic and lexical complexity on third-grade students' comprehension of science texts. A total of 16 expository texts were designed to represent systematic differences in levels of syntactic and lexical complexity across four science-related topics (*Tree Frogs, Soil, Jelly Beans and Toothpaste*). A Latin-square design was used to counterbalance the order of administration of these 16 texts. After reading each text, students responded to a post-test comprehension measure (without access to the text). External measures of reading achievement and prior vocabulary knowledge were also gathered to serve as control variables. Findings show that lexical complexity had a significant impact on students' comprehension on two of the four topics. Comprehension performance was not influenced by the syntactic complexity of texts, regardless of topic. Further, no additional effects were found for English language learners. Potentially moderating and confounding issues, such as the inference demand of syntactically simple texts and the role of topic familiarity, are discussed in order to explain the inconsistency of the findings across topics.

Keywords: Text complexity, reading comprehension, science literacy

Introduction

Recently, scholars have highlighted the need for increased attention to informational texts in elementary schools, especially primary-level classrooms (Donovan & Smolkin, 2001; Duke, 2000). The argument for this shift in textual diet is that increased attention to informational texts will improve many of the things that matter in students' later development: world

* Diana J. Arya, Norwegian Center for Science Education at the University of Oslo, Norway.
E-mail: d.j.arya@naturfagsenteret.no

ISSN:1307-9298

www.iejee.com

knowledge, monitoring and problem-solving strategies, and dispositions toward academic reading. While all disciplines have benefited from this shift in emphasis, science has received the most attention. For example, in 2010, an entire special issue of the Journal, *Science*, was devoted to the literacy-science interface—a remarkable departure for a public access journal that normally focuses on research and policy for the hard sciences. Science requires numerous firsthand experiences; however, appropriate texts can have a critical role in science learning (Cervetti & Barber, 2009; Guthrie, McRae & Klauda, 2007). Science texts provide readers with a purpose for reading and additional exposure to key science concepts that lead to deeper conceptual understanding (Guthrie, Anderson, Alao, & Rinehart, 1999; Palincsar & Magnusson, 2001; Romance & Vitale, 1992; 2006).

Although the academic benefits of science texts are evident, they pose challenges to teaching and learning. In particular, the vocabulary of science texts can be dense and complex (Armstrong & Collier, 1990; Schleppegrell, 2004; Snow, 2010). Elementary science texts have been criticized for being inaccessible because they introduce the reader to many unfamiliar words yet fail to explain them in ways that connect with students' experiences (Armbruster, 1993; Armstrong & Collier, 1990; Norris & Phillips, 2003; Rutherford, 1991). One of the benefits of having a science text is to help clarify and extend scientific concepts that students encounter during firsthand investigations (Duke & Bennett-Armistead, 2003; Donovan & Smolkin, 2001). However, for young students who are still developing literacy skills, as well as academic vocabulary, science texts containing unfamiliar terms can be very difficult to comprehend.

In the development of student science texts, there is a tension between conceptual explicitness (which often requires more complex syntactical realizations and rare, concept-oriented vocabulary words) and linguistic simplicity (which generally requires less complex syntactic realizations and simpler vocabulary). Matters of syntactic complexity were salient in the text comprehension research of the 1970s and even into the early 1980s (Pearson & Camparell, 1981), but text structure yielded to other emphases, most notably comprehension strategy work, in the 1990s and early 2000s (see Pearson, 2009). The need to re-examine these factors is greater than ever in light of two recent developments. First the dramatic increase in the numbers of students with diverse linguistic and cultural backgrounds (US Census, 2000) and the challenges many linguistically diverse students experience on tasks such as the NAEP Science assessment (Gutierrez & Rogoff, 2003; Lee, & Luykx, 2005; Shaw, 1997) requires us to take a closer look at text features that may prove especially challenging or supportive for English language learners. A major challenge is identifying text features that make information more accessible for ELLs. Second, the advent of the new Common Core State Standards in English language arts (CCSS, 2010) has upped the ante on standards of text complexity; educators are being challenged by these standards to increase the complexity of texts students read at every grade level by at least a half grade in measured readability. This means that all students are going to be asked to read texts with more complex syntax and more difficult vocabulary. In this study, we investigate the extent to which lexically and syntactically complex realizations of content hinder or help comprehension—and whether these two factors interact to provide either unique scaffolds or barriers to acquiring important science concepts.

Gauging Text Complexity

Syntactic Complexity. The "simple" view of syntactic complexity evident in readability formulas such as the *Flesch Ease of Reading* formula (Flesch, 1948) holds that the fewer words in a sentence, the less difficult it is for readers to comprehend (Klare, 1984). This perspective, however, may be misleading; more words may simply be an alias for more ideas or, even more likely, more complex ideas. In other words, conceptual complexity could be driving difficulty, and the number of words in a sentence simply indexes (rather than causes) that complexity. In psychological terms, the longer the sentence, the greater the likelihood that multiple discrete ideas, called propositions, are embedded in it (Kintsch, 1998). Examples 1-3 illustrate this point.

1. Tree frogs have red eyes.
2. Tree frogs have red eyes that help them see and find food.
3. Tree frogs have red eyes that help them see and find food at night.

Example 1 conveys two complete ideas or propositions: (a) tree frogs have eyes, and (b) these eyes are red. Example 2 has two additional propositions: these red eyes help the frog to (a) *see* food and (b) to *find* food. Example 3 actually adds three more propositions, one explicit and two implicit. The explicit proposition is that the frogs do the seeing and finding at night. That proposition invokes two more entailments: that (a) that frogs are *awake* at night and (b) that their eyes help them to see *in the dark*. The amount and explicitness of the information provided in each sentence increases as the number of embedded structures (e.g., adjectives, relative clauses, and prepositional phrases) increases. Readers must be able to unpack the propositions within complex sentences and establish their logical relations to one another to understand all of the information presented.

Any account of text difficulty that uses sentence length to establish the readability of texts assumes, at least implicitly, that unpacking the propositions within a complex sentence is more difficult than making connections across related propositions stated in simple sentences. A short sentence in itself may be easier to comprehend than a complex one. However, the challenge may come when the reader needs to construct a cohesion model of meaning from a series of short sentences. To illustrate this distinction, a complex sentence such as the third example above can be broken up into five simple sentences, as in Example 4:

4. Tree frogs have eyes. These eyes are red. These eyes help them see. They help them find food. The tree frogs are awake at night.

Just as the complex sentence required readers to unpack propositions within the sentence, having to connect ideas across discrete, simple sentences may place other task demands on readers. Connective cues (e.g., conjunctives, conjunctive adverbs, and relative clauses) and other embedded structures serve as markers to guide readers to a full understanding of the ideas presented. Eliminating these connective cues may increase the inference burden on readers (Bowey, 1986; Pearson & Camperell, 1981); relationships, such as cause-effect or problem-solution, or sequence, that were explicitly cued in the more complex versions have to be inferred in the less complex versions. Ozuru, Dempsy, Sayroo and McNamara (2005) found that adding cohesive devices such as connectives that made relationships between sentences more explicit were beneficial for students reading science texts about unfamiliar topics. Students were able to correctly answer more questions when texts had syntactic structures that made meaning more explicit than when texts were less cohesive. Similarly, Rawson (2004) found that texts that presented more ambiguous syntactic structures with unmarked, reduced relative clauses (*the girls told about the movie were excited*) were more difficult for college students (all of whom had high reading abilities) than texts with more explicit structures, containing marked clauses (*the girls who were told about the movie were excited).*

While some complexity in sentences can support readers in comprehending text, presumably there is a point where sentences can become too complex for novice or inexperienced readers. Several factors likely influence where this tipping point occurs. Developmental level and reader proficiency appear to be two such factors in that older readers and more proficient readers demonstrate greater comprehension of grammatically complex structures than younger and less proficient readers (Nation & Snowling, 2000; Willows & Ryan, 1986).

Background knowledge or conceptual familiarity of the topic may also influence readers' abilities to comprehend the embedded structures of complex sentences. Goldman and Bisanz (2002) reported that novice and less proficient readers who did not have background knowledge of a topic were less able than more knowledgeable or more proficient readers to avail themselves of embedded structural cues. However, McNamara, Kintsch, Songer, and Kintsch (1996) found that science texts containing greater number of embedded structures that clarify or highlight information (e.g. use of connectives or embedded explanations) benefit readers with less knowledge of the concepts whereas texts with "cohesive gaps" (i.e., fewer connectives and embedded clauses) that require students to make inferences about relationships and concepts benefit readers with a strong level of prior knowledge (McNamara, 2001; McNamara et al., 1996); in short, less knowledgeable readers are aided by the scaffolding of explicit cues but more knowledgeable readers are aided by the challenge of a text that needs "fixing". The current study attempts to address this conflict. Thus when it comes to the issue of syntactic complexity, a trade-off may well exist: What is made more easily accessible by complexity (seeing the relations among propositions) is made more esoteric by simplicity. What is made easier to comprehend by simplicity (getting unitary ideas through the veil of working memory) is rendered complex by the addition of embedded structures.

Lexical Complexity. Sentence length, typically an alias for syntactic complexity, is often, indeed almost universally, coupled with vocabulary difficulty in readability formulas in order to determine overall accessibility of a given text (Flesch, 1948, 1979; Lennon & Burdick, 2004). Vocabulary difficulty is generally indexed by how frequently a particular word generally appears in texts (Zeno, Ivens, Millard, & Duvvuri, 1995). The assumption is that the more exposures a reader has to a particular word, the more a reader learns about it and, in turn, the more accessible that word (and the message in which it is embedded) becomes. Indicators of familiarity have long been used to estimate the readability of text (e.g., Cunningham & Stanovich, 1998; Snow & Sweet, 2003; Stahl, 1999).

Word frequency is strongly correlated with word knowledge, which is a crucial aspect of reading comprehension (NICHD, 2000; RAND Reading Study Group, 2002); simply put, the more frequently a word occurs in a language the greater the likelihood that students will know its meaning. Research on vocabulary suggests that texts containing few unknown words provide readers with an appropriate source from which to develop fluency and word knowledge (Beck & McKeown, 1991; Qian, 2002; Vellutino, 2003). Thus, the more students read texts with very few rare words, the greater their chances in developing a solid understanding of the unfamiliar concepts that are present, allowing them to comprehend texts with additional lexical complexity in the future (Nagy & Scott, 2000; Stanovich, 2000). However, it may be argued that the more frequent a word, the greater the likelihood that, while students will know its meaning, its meaning may be less precise (Carey, 1985; Gopnik, 1996). This may especially be so with words in science where a less frequent word such as *astronaut* conveys a level of precision that a generic word like *man* does not. Conversely, too many unfamiliar or complex vocabulary words within science texts may inhibit readers' ability to learn concepts through reading (Shymansky, Yore, & Good, 1991; Stahl, 1999). We expect students to infer word meanings from context; it is a required part of skilled, strategic reading. However, if there are too many unknown words in the surrounding context, there may be no meaning base from which a student could infer the meaning of a particular word. Contrast the challenge of inferring the meaning of *habitat* in Examples X and Y:

X. The soil in the alluvial plane, rich in nutrients and decomposers, provided an optimal habitat for our earthworms.
Y. The soil along the river provided a good habitat for our earthworms.

Science texts are purported to have more than twice the number of rare words as texts from any other discipline, thus creating a vexing challenge for developers of science literacy curricula: How can they create considerate and accessible texts for young readers that also do justice to the concepts students are supposed to acquire (Hayes & Ahrens, 1988)? Just as with syntactic complexity, there is a potential trade off in lexical complexity. Rare words have a level of precision that high frequency words do not. However, the presence of too many rare words may make a text inaccessible to readers.

Vocabulary familiarity (complexity) has a direct relationship to readers' knowledge about the topic, which has a great impact on comprehension (Kintsch, 1998; RAND Reading Study Group, 2002; Smagorinsky, 2001; Snow & Sweet, 2003; Stahl, 1999). As one becomes more familiar and experienced with a topic, knowledge of contextualized meanings of words develops as well (Anderson & Freebody, 1981; Kintsch, 1998). In experiments that use association and priming tasks, skilled readers have been found to approach a text with an organized network of knowledge called schemata. These allow readers to integrate new information with prior knowledge (Kintsch, 1998; RAND Reading Study Group, 2002; Smagorinsky, 2001; Snow & Sweet, 2003) and, in the process, enhance their schemata even more. The stronger one's prior knowledge about a particular subject, the greater one's ability to read and comprehend texts quickly and efficiently (Kintsch, 1998). The connections that readers make with text are dependent on their knowledge base and ability to retrieve the most relevant meaning from alternatives in their mental lexicons (Kintsch, 1998; Smagorinsky, 2001; Wilson & Sperber, 1987).

Just as students' prior knowledge about particular concepts facilitates comprehension, a lack of knowledge about concepts within a text can have a detrimental impact on understanding. Bailey (2007) conducted a language analysis of American standardized achievement tests and found that academic language (i.e., words often used in tests such as *examine* or *cause*) confounds the ability of English Language Learners (ELLs) to demonstrate their understanding of the construct that is being assessed in English. Similarly, Droop and Verhoeven (1998) found in their study of third grade students learning Dutch as a first or second language that lexical complexity (defined in terms of word frequency) as well as cultural relevance impacts text comprehension. However neither of these studies examined the impact of syntactic complexity or its interaction with lexical complexity in academic language.

The Current Study

The aim of the present investigation was to compare the effects of syntactic and lexical complexity on students' understanding of science content. Students' comprehension of texts was examined as a function of two dimensions of syntactic complexity (simple, complex) and two dimensions of lexical complexity (simple, complex); additionally, the main and interaction effects of syntactic and lexical complexity were examined through the lenses of reading ability and prior knowledge.

Language status was also considered as a potential confounding factor on the comprehension of these texts. Text accessibility is an important issue for ELLs because they must have the opportunity to read extensively in texts at their level of reading ability in order to improve comprehension and fluency (Cunningham & Stanovich, 1998; Elley, 1996; Grabe, 1991; Snowling & Nation, 1997). However, few studies of readability have investigated the effects of text difficulty on the comprehension of ELLs. Further, no such study has focused on both lexical and syntactic complexity while holding issues of cultural relevance constant.

Specifically, the following questions are addressed in this investigation:

1. Do syntactic and lexical complexity affect comprehension of science texts for third graders?
2. How do these two forms of complexity interact to produce unique combination effects on comprehension?
3. Are there any additional effects of syntactic and lexical complexity for ELLs?

We anticipated that, the greater the complexity of a given science text (as measured by embedded clauses and difficult vocabulary), the more skilled a reader must be to successfully understand the text. Thus, scores on general reading assessments such as informal reading inventories and state tests should predict scores on an assessment of comprehension of science content. Based on a long history of readability research, we also hypothesized that lexical complexity might have a greater impact on performance than syntactic complexity, but that the interaction of syntactic and lexical complexity would have the most debilitating effect on comprehension. In short, only the very best readers as defined by reading test scores would be able to handle the difficulty imposed by texts that are complex on both syntactic and lexical criteria.

Questions about the manner in which prior knowledge of words and lexical complexity influence students' comprehension and how these constructs contrast with syntactic complexity merit particular attention with science texts. It is possible that, when complex ideas are communicated with accessible (i.e., high frequency) vocabulary, syntactic complexity does not matter as much as it does when technical (low frequency) vocabulary is used. Further, a reader's experience with certain subject matter may determine the degree to which lexical complexity, syntactic complexity, or both affect understanding.

Method

Participants

This study included all 142 third-graders who had returned parental consent in 10 classrooms in four non-charter public schools. According to California state regulations operating at the time of the data collection, no K-3 classroom could enroll more than 20 students. An average of 14.3 students per classroom (minimum = 12, maximum = 19), which was approximately 75% of total third grade enrollment across these schools, participated in the study. All four schools were located in northern California and varied according to urbanicity, ethnicity, and percentage of ELLs, defined by language spoken at home. The reason for this ELL distinction is that three of the participating schools had no English language program, thus having no school language designation for students who are learning English as a second language. This information was reported by the participants and confirmed by the parental consent letters. Students who spoke only English at home were not considered to be ELL. All other students (49 students, 34% of total participants) were considered to be ELL. Of the 49 ELL students, 23 (16%) spoke only Spanish at home while 11 (8%) spoke a mix of English and Spanish. The remaining 15 students (11%) spoke one of various Asian or European languages at home.

Two of the four participating schools (3 classrooms total) were urban, while one was located within a suburban area (one classroom) and one is rural (6 classrooms total). The four schools ranged in percentage of ethnic minority (i.e., other than White, 44%--73%) as well as language minority (other than English, 12%--59%) students. Half of the total population of participants were Caucasian (71 students) while the other half represented Hispanic (44 students, 30% of total participants), African American (15 students, 10% of total participants), and Asian (eight students, 6% of total participants) ethnicities with a remaining few (five students, 4% of total participants) representing other or mixed ethnicities. Note that SES was not reported for this study: as of 2005, it is illegal to access this particular demographic information according to California state regulations, even for classroom teachers, regardless of consent or approval of the university's internal review board.

In the initial phase of the study, each participant read a narrative passage orally from the *Qualitative Reading Inventory (QRI)* (Leslie & Caldwell, 2000) and answered questions about it. Students who were not able to read at least 75% of the narrative text (five participants total) were excused from continuing on with the assessment procedure in order to prevent potential frustrations. Thus, a total of 142 participants continued with the study. Participants demonstrated a broad range of abilities in fluency (words read within a minute, WPM) and comprehension (number of correct responses to questions about the passage reading) on the *QRI*. The mean WPM performance was 97 with a standard deviation of 36. The mean comprehension score was 5.8 (based on a total of 8) with a standard deviation of 1.7.

Assessments

Assessments were administered across three sessions that spanned a three-week period. In the first session, the QRI and a measure of students' knowledge on the specific topics of the experimental texts were administered. In the second and third sessions occurring three weeks later, students were given the experimental texts. Two additional measures of student reading achievement were obtained: (a) teachers' ranking of student reading proficiency and (b) student scores on the Standardized Testing and Reporting Program (STAR) (California Department of Education, 2007) from the previous spring. The mean score of the STAR was 351 with a standard deviation of 62. All assessments and scores described here were used to establish a baseline of reading abilities for all participants.

Qualitative Reading Inventory (QRI). As already described, students individually read a third-grade, narrative passage of the *QRI* and answered explicit and inferential questions about it (Leslie & Caldwell, 2000). A student's oral reading of the text was timed and miscues recorded. The oral readings and responses to comprehension questions were tape-recorded to establish fidelity of different investigators' on-the-spot recording of miscues. The authors of the QRI report very high alternate form reliability ($r = .9$) as well as high correlation with an unidentified standardized reading test ($r = .7$). This form of assessment was used not only for its reliability, but also for the fact that participating teachers and students are familiar with this more qualitative format of assessment. Thus, the teachers could also use student performance on the QRI formatively for general educational purposes.

Prior vocabulary knowledge. A prior knowledge measure was developed by identifying six words for each of the four topics that were the focus of the experimental portion of the study: tree frogs, toothpaste, jelly beans, and soil. Sixteen of the words in the 24-item measure represented highlighted science concepts in the lexically academic forms of the experimental texts (four items per topic). All words were within the same general range of frequency, from 46 to 53 on the SFI index (Zeno et al., 1995). The remaining eight items consisted of either words representing science concepts that were not part of the experimental texts (e.g., *terrarium)* and or cross-disciplinary words (e.g., *determine*) that were not in the experimental texts but were within the same range of frequency. This second group of words was included in order to obtain a measure of general lexicon as well as of the specific topics in the study.

The 24 words were then randomly organized into six groups. A student's task was to match a word with its definition for each group of words. Definitions were short, everyday descriptions of the words, such as *to dig* for *burrow*. This task was not timed and was completed in small, investigator-supervised groups.

Teacher rankings of students. Teachers were asked to rank students, beginning with 1 (the strongest reader in the classroom). Teachers completed these rankings without receiving feedback on students' performances on the *QRI* or the prior knowledge measure.

State assessment. Students' performances on the state's Standardized Testing and Reporting Program (STAR) (California Department of Education, 2007) from the end of the prior academic year were obtained as an external measure. This measure was used as a covariate with the *QRI* to establish external validity of the vocabulary pre-assessment and comprehension measure of the experimental texts. Forty-two scores were missing due to record unavailability (15 total) and missing teacher files (two of the six classroom teachers within the rural school were unable to locate scores).

Experimental texts. Sixteen texts of approximately 200 words in length were written, with four versions of each of four different science topics. Topics were identified from the national science education standards (NRC, 2001) to represent the three strands of life, earth, and physical science.

The creation of the experimental texts began with a single text for each topic. This initial text had three sections: (a) an introductory section of 50 words that was common across all conditions, (b) a manipulated section of approximately 100 words (within a 9-word range) that differed according to condition, and (c) a concluding section of 50 words that was common to all conditions. The introductory and concluding sections of the text used "simple" syntactic forms and "everyday" lexical content.

Table 1. *Example of Manipulated Version of Topical Texts: Jelly Beans*

	Everyday Vocabulary	*Academic Vocabulary*
Syntactically Simple	A scientist wanted to make a new flavor. He wanted to make grass flavor. Grass is not safe food. He could not use real grass. He used other things. These things are safe. His new jelly bean smells like grass. It tastes like grass.	One scientist wanted to **invent**[1] a flavor. This was grass flavor. Grass is not **edible.** He could not **manufacture** the flavor. He used different **ingredients.** This jelly bean had the **odor** of grass. It had the taste of grass.
Syntactically Simple	A scientist wanted to make a new flavor. He wanted to make grass flavor. Grass is not safe food. He could not use real grass. He used other things. These things are safe. His new jelly bean smells like grass. It tastes like grass.	One scientist wanted to **invent**[1] a flavor. This was grass flavor. Grass is not **edible.** He could not **manufacture** the flavor. He used different **ingredients.** This jelly bean had the **odor** of grass. It had the taste of grass.
Syntactically Embedded	One scientist wanted to make a new flavor, grass flavor, *by*[2] using other things *because* grass is not safe to eat. He could not use real grass to make the flavor, *but* it smelled *and* tasted like grass.	One scientist wanted to invent a flavor, grass flavor. by using different ingredients because grass is not edible. He could not use grass to manufacture the flavor, but this jelly bean had the odor and taste of grass.

[1]Academic vocabulary
[2]Word indicating an embedded structure

Table 2. *Indexed Features of Syntactic and Lexical Complexity*

	Syntactic Complexity (Average Number of Propositions within Version)		*Lexical Complexity (Average Standard Frequency Index (SFI) within Version)*	
	Simple	Embedded	Everyday	Academic
Tree Frogs	2.9	7.3	65.9	52.1
Soil	2.9	7.3	63	48.3
Jelly Beans	2.6	7	67	47.5
Toothpaste	2.7	6.8	63.4	47

The middle section of the text was rewritten so that there were four texts for each topic: (a) syntactically simple with everyday vocabulary (simple/everyday), (b) syntactically complex with everyday vocabulary (embedded/everyday), (c) syntactically simple with academic vocabulary (simple/academic), and (d) syntactically complex with academic vocabulary (embedded/academic).

For this study, a high level of syntactic complexity was defined as the presence of two or more embedded structures within a sentence; sentences with one or no embedded structures were deemed as low in syntactic complexity. Embedded structures included relative clauses, nominalizations, appositives and multiple modifiers. An illustration of the "treated" portion of a text and the nature of embedded structures appears in Table 1.

A propositional analysis (Kintsch, 1998) was used to determine the difference between syntactically simple and complex texts. The average number of propositions per sentence for the simple and embedded versions of the texts is summarized in Table 2. Across the four topics, the difference between the simple and complex version is consistently about 4 propositions per topic.

Lexical complexity was indexed by the presence or absence of academic (cross-disciplinary or scientific) words that directly relate to science concepts or processes and are beyond the 1000 most frequent words according to Zeno et al. (1995). High-frequency words (words within 1000 most frequent) are referred to as "everyday words." To verify the differences across these passages, the standard frequency index (SFI) of the words in each passage were computed. The higher the SFI, the more frequently the word is used in texts (e.g., *the* = 88.3; *sanitize* = 25.6). The average SFIs for academic and everyday versions across the four topics are reported above in Table 2. Averaged across the four topics, the difference between the mean SFI values of the academic and everyday versions was 16 (the difference between the mean SFI value of each of the individual topics were within three points of this). Since the remaining portions of the texts (i.e., the first and last 25% of each text) are equivalent, and since the function words for all manipulated versions are high in frequency, the focus of the analyses was on the everyday and complex version of academic words.

For each topic, 10 questions were constructed to measure students' comprehension. Half of the questions were multiple-choice and half required short-answer responses. An example of a multiple-choice question is the following: *What makes plants grow? a. rocks; b. vitamins; c. bugs; d. wind.* The short-answer responses were constructed to elicit a specific response, such as the following: *Write two ways that animals help plants*. The questions for a given topic were the same, regardless of the manipulated condition that students received. Four of the 10 questions targeted the content of the manipulated portion of the text; the remaining six questions referred to the first and last 25% portion of the text (three questions for each portion). Two of the four questions for the treated portion were explicit recall of information from the text and two required the student to make inferences based on what they read from this portion. The remaining six questions also consisted of both direct recall and inferential questions.

The short-answer questions (e.g., *How do frogs get away from their enemies?*) were scored on a scale of 0-1-2. A rubric was constructed to assign no, partial or full credit. No credit was given to responses that were irrelevant (e.g., *they like to swim*). Partial credit was given to responses that included part of the intended answer (e.g., *they hop around*). Full credit was given to complete and accurate answers (e.g., *they hop around really fast*). A sample of 20% of the responses was double-scored; the inter-rater agreement was 95%.

Reliability and validity of measure

All experimenter-designed assessments were piloted to determine validity and reliability. After revision, the prior knowledge assessment had a Cronbach's alpha coefficient of .85 and correlated strongly with the QRI timed miscue measure (.65, $p < .01$), teacher ranking of reading ability (.57, $p < .01$) and performance on the STAR (.67, $p < .01$).

The comprehension assessments for the experimental texts on the four topics, *Tree Frogs, Soil, Jelly Beans* and *Toothpaste,* had a Cronbach's alpha coefficient of .86. These comprehension assessments strongly correlated with the state reading assessment (.56, .67, .74, .63; $p < .01$) and the QRI timed miscue measure (.51, .50, .51, .51; $p < .01$).

Procedures

Three experienced researchers collected all of the data for the present study. To reduce the possibility of priming the participants on key vocabulary, the prior knowledge measure was administered individually in one session, along with the QRI task, at least three weeks prior to the experimental reading task. The passage reading/comprehension tasks took place in two sessions as whole-class events on two separate days; each of these sessions lasted approximately 50 minutes.

All participants read four passages with the constraint that each student received each topic and each version once and only once. There were 4 topics and 4 versions per topic, yielding 16 unique reading tasks (a passage followed by the comprehension items connected with that particular topic). These reading tasks were assigned to participants using a Latin-square design, which resulted in complete counterbalancing for the order in which both topic and version were presented. In other words, each of the 16 reading tasks was completed equally often in the first through fourth testing positions across students. To avoid fatigue, participants completed two reading tasks on the first day and two on the second day of testing. As an example, one student might have read *Tree Frogs* in the syntactically simple/everyday vocabulary version and *Soil* in the syntactically simple/academy vocabulary version on day one, followed by *Toothpaste* in the syntactically complex/everyday vocabulary version and *Jelly Beans* in the syntactically complex/academically vocabulary version on day two. It required a total of 64 participants to complete one complete replicate of the 4 topics X 4 versions X four serial testing positions design.

Participants were given as much time as needed to read the text and then answer the questions, but each text was collected directly before distributing questions. They were required to answer each set of questions based on memory of what had been read, without the opportunity to look back at the text. Tables 3a and 3b show the total performance on each text by version and topic as well as specific performance on only the treated portions.

Results

A series of 2-step (students were level 1 and classrooms, level 2), hierarchical linear models were fit to the data to examine the relationship between treatment (syntactic and/or lexical complexity) and performance on the treated sections of the text, while simultaneously accounting for variance due to the clustering of students within classrooms. A random intercept was included in the model; it permitted different mean performance levels across classrooms. No random slopes were included in this model due to the small number of classrooms ($N = 10$) as well as the implausibility and irrelevance of classroom-specific effects of treatment on performance. No additional classroom variables were considered in the present study. Such analyses, which would have allowed for more level-2 covariates, would have required a much larger sample of classrooms than was available.

Error-variance histograms revealed that the error variance from each of the regression models fit was normally distributed. Also, predicted-versus-observed scatterplots of the outcome variables revealed that the error variance was constant across the range of data. Thus, the assumptions of regression modeling were met for the data used in this study.

This study uses a modest form of HLM, with a random intercept only and no level-2 covariates. In Raudenbush and Bryk's (2002) notation, our full model (which corresponds to Model 3 described below) is described by this formula:

Table 3a. *Means and SDs for Total performance on Designed Texts*

Topic→	*Tree Frogs*	*Soil*	*Jelly Beans*	*Toothpaste*
Version 1 (simple/everyday)	10.6 (3.3)	9.5 (3.3)	6.2 (3.6)	9.7 (3.2)
Version 2 (complex/everyday)	10.7 (3.4)	8.7 (3.1)	5.1 (3.6)	9.7 (3.2)
Version 3 (simple/academic)	10 (3.2)	7.5 (3)	6.3 (2.8)	10.2 (3.7)
Version 4 (complex/academic)	9.6 (3.1)	7.3 (2.6)	6.1 (2.9)	9.6 (3.1)

Table 3b. *Means and SDs for Treated Portion of Designed Texts*

Topic→	*Tree Frogs*	*Soil*	*Jelly Beans*	*Toothpaste*
Version 1 (simple/everyday)	4.4 (1.7)	3.4 (1.6)	2.2 (1.8)	3.4 (1.4)
Version 2 (complex/everyday)	4.5 (1.6)	3.2 (1.5)	1.8 (1.6)	3.5 (1.2)
Version 3 (simple/academic)	4.0 (1.7)	2.6 (1.7)	2.1 (1.5)	3.9 (1.6)
Version 4 (complex/academic)	3.5 (1.7)	2.6 (1.6)	2.2 (1.5)	3.5 (1.3)

In the present study, the model form described above was fit four times, once for each of the four topics. Although the multiple models were fit using the same participants, a Bonferonni-like correction was not applied in this situation given that the same question was asked four times, once for each topic. Naturally, we hoped that results from the four model sets would converge.

The first model fit (Model 1) is a variance-components model with no covariates and is presented to illustrate the amount of total variance in performance that can be attributed to classroom-level effects. Model 2 adds the control variables, and Model 3 adds the independent variables. Since the interaction between syntactic and lexical complexity was not significant, it was dropped for the final model (Model 4). This variance components model indicates that a significant amount (6.4%, $p < .05$) of variation in performance is between-classrooms. Since the various text conditions were assigned randomly to students within classrooms, it was important to control for classroom-level effects in order to accurately assess treatment differences within all ten classrooms included in the analysis.

Model 2 adds in the covariates, which are home language (i.e., ELL status) and four pretest scores (STAR from grade 2, prior vocabulary knowledge, and the fluency and comprehension scores for the 3rd grade QRI passage). Pretest scores were a highly significant predictor of performance; ELL status was not, after controlling for pretest scores. Thus, ELL status did not explain any additional variance in performance on the designed texts. The random intercept variance remained significant, but its share of the variance was reduced greatly in comparison to Model 1, indicating that much of the variance between classrooms is attributable to student background characteristics and prior achievement.

Model 3 adds in the independent variables: presence of syntactic complexity, presence of lexical complexity, and an interaction term between the two. All three of these variables were non-significant. We then dropped the interaction term from the model, leaving model 4, in which lexical complexity affected performance but syntactic complexity did not.

Model 4 explains a significant amount of variance for only two of the four topics, *Tree Frogs* and *Soil*. Similar results were not obtained for *Jelly Beans* and *Toothpaste*; for the latter two topics, neither lexical nor syntactic complexity affected performance.

This final model suggests that high lexical complexity (i.e., more low frequency words) in the text is associated with lower performance on the test ($p < .05$). As would be predicted by the design of the passages, the impact of lexical complexity was limited to items in the middle 50% of the passage (the manipulated portions); lexical complexity did not explain any significant portion of variance in responses for comprehension items relating to the first and final sections of the texts. A model with only syntactic complexity as a predictor variable was also fit to the data, but was not significant at the 0.05 level. Model 4, with lexical complexity predicting comprehension differences across forms, is presented in Table 4 for all four topics.

These inconsistent results prompted a series of post-hoc investigations into the particular conditions under which lexical complexity of a text may affect comprehension of that text. The most obvious candidate to explain the inconsistent patterns is background knowledge of particular concepts across the four topics. The knowledge of concepts explanation was explored in two ways. The first was an examination of the SFI indices of frequency from the Zeno et al.'s (1995) corpus; these data appear in Table 2. Differences between the SFIs for the academic and everyday versions of the texts for the four topics were calculated. The observed average SFI differences between levels of lexical complexity, which were (in order of magnitude), *Jelly Beans*: 19.5; *Toothpaste*: 16.8; *Soil*: 14.7; and *Tree Frogs*: 13.8, would have predicted the greatest between-version differences in comprehension on the *Jelly beans* and *Toothpaste* passages. Ironically, just the opposite pattern was evident in the data, with the greatest differences between academic and everyday versions on *Soil* and *Tree Frogs*, the two topics with the smallest differences between the everyday and academic versions. Thus, SFI index does not provide a suitable explanation for the apparent interaction between topic and lexical complexity.

The second way in which background knowledge was considered was to examine the relationship of the prior knowledge vocabulary measure to comprehension of the topics. Recall that the prior knowledge vocabulary measure correlated strongly with students' comprehension of the manipulated portions of the texts: *Tree Frogs*: .52; *Soil*: .59; *Jelly Beans*: .65; *toothpaste*: .67 ($p < .01$). The mean scores (out of a maximum of 4) and standard deviations of the prior vocabulary assessment items for the four topics are as follows: *Tree Frogs:* 2.3 (*sd*, 1.3); *Soil*: 1.7 (*sd*, 1.3); *Jelly Beans*: 2.9 (*sd*, 1.1); *Toothpaste*: 2.8 (*sd*, 1.0). When the simple effects were calculated across these four means, the analysis showed that "academic vocabulary" used to create the complex versions of the passages yielded significantly different pre-test vocabulary results across the four topics. The pre-test academic vocabulary performances for *Toothpaste* and *Jelly Beans*, which did not differ from one another, were significantly easier than either *Soil* or *Tree Frogs*; additionally, *Tree Frogs* was easier than *Soil* ($p < .01$, in all cases); in sum: (*Jelly Beans= Toothpaste*) > (*Tree Frogs* > *Soil*). Thus, the empirical measure of students' prior knowledge of words was a more accurate predictor of lexical complexity than the SFI index. It is the only plausible explanation of the differential effect of lexical complexity across topics.

Discussion

The present study was designed to address the question of whether lexical or syntactic factors exert greater influence on the comprehension of elementary science texts. Based on previous research on text accessibility, it was expected that syntactic and lexical complexity would each affect students' performance on science texts, and that these two types of text complexity together would additionally impact student performance. In order to test this hypothesis, 16 texts that varied in syntactic and lexical complexity across four different topics were constructed. Students read texts that ranged in complexity, each from a different topic.

Contrary to our hypotheses, syntactic complexity did not explain variance in performance across any of the four topics. It is difficult to interpret our results on syntactic complexity. As established in the review of research on this topic, opinions are divided as to whether or not explicitness, as defined by embedded clauses and connective cues, hinders or aids comprehension. It is possible that different sorts of cognitive loads effectively canceled out differences between the syntactically simple and complex versions: our syntactically simple versions required students to engage in a great deal of inferencing to create the logical links between sentences (e.g., A caused B or A happened before B). By contrast, the syntactically complex versions required readers to hold many embedded constructions and cues in short term memory to unpack those logical links. However, since reading ability (as measured by the QRI and STAR test) and the prior knowledge assessment did not interact with syntactic complexity, it is difficult to sort out what was happening across levels of syntactic complexity. We certainly were not able to replicate the McNamara et al (1996) finding of an interaction between students' level of prior knowledge and the cohesion of the texts as indexed by strong use of cohesive ties between clauses and sentences. Future studies might include gradations of syntactic complexity in order to begin to unpack this mystery. The other possibility is that the methodology used for measuring comprehension obscured the real impact of syntax. It may be that syntax achieves its effect on comprehension in the "search" process readers engage in when they consult the text to find exact answers to explicit questions or clues to help them draw inferences. By taking away the texts during the comprehension assessment, we may have pre-empted the very mechanism (text search) through which syntactic explicitness achieves its effect.

Table 4. *Regression Results without Interaction Terms (Model 4)*

Predictor	Topic 1 (*Tree frogs*)	Topic 2 (*Soil)*	Topic 3 (*Jelly Beans*)	Topic 4 (*Toothpaste*)
Intercept	1.82** (.52)	1.23** (.34)	.80 (.53)	3.77** (.64)
Home language	-.47 (.34)	-.46 .(23)	-.53 (.35)	-.61 (.47)
QRI (pretest)	.46** (.07)	.27** (.05)	.24** (.08)	.54** (.10)
Syntactic complexity	.06 (.25)	-.22 (.17)	.06 (.26)	-.46 (.35)
Lexical complexity	-.55* (.24)	-.54** (.16)	.13 (.25)	.36 (.34)
Variance component of:				
Classroom mean, u_{ij}	.044*	.185*	.086*	.228*
Level-1 effect, r_{ij}	1.999	.854	2.103	3.599

Note: The results represent a set of non-nested multilevel models, fit to the same participants using different topics. Standard errors are given in parentheses.
* p < .05; ** p < .01.

Lexical complexity significantly influenced comprehension performance for texts on two of the four topics, *Tree Frogs* and *Soil*, but not for texts on *Jelly Beans* and *Toothpaste*. This finding was consistent across all participant groups, including ELLs. A possible explanation is that prior knowledge of vocabulary, rather than any established index of word frequency, determines how difficult a lexically complex text will be for a student. Although, for example, *bacteria* is considered a very low frequency word, 62% of the participants were able to correctly identify its meaning. Further, *essential*, a word with a comparable SFI value to *bacteria* (SFI=56) was a much less familiar word, at least for our sample of students, in that less than half (42%) of the students were able to correctly identify its meaning. Assuming that world and word knowledge is shaped by experience, it is plausible to assume that most eight year olds (the average age of our sample) have visited the dentist several times and have learned about dental hygiene, including words such as *bacteria*. The role of conceptual familiarity as a predictor of text comprehension has been commented upon in previous research (Cunningham & Stanovich, 1998; Kintsch, 1998; Smagorinsky, 2001; Snow & Sweet, 2003; Stahl, 1999), thus giving strength to this admittedly speculative explanation for the interaction between topic familiarity and lexical complexity. However, it is important to note that our explanation of the inconsistent lexical complexity effect are tentative at best and require further investigation. Future studies on the effects of lexical complexity should include measures of students' prior knowledge in order to assess conceptual familiarity adequately.

A specific interest in the present study was the effect of variations in text complexity on the comprehension of ELLs. Language status did not explain any additional variance in performance beyond the general findings in this study. Thus, lexical complexity was the only significant factor in comprehension performance for ELLs. This finding is consistent with research by Proctor, August, Carlo and Snow (2005) who reported that L2 vocabulary knowledge was a significant predictor of L2 text comprehension of ELLs. Our findings did not reveal any significant differences in comprehension performance between native English speakers and ELLs, thus suggesting a global model of comprehension, seemingly contrary to Proctor et al.'s (2005) conclusion that we need an L2-only model of comprehension. However, due to differences in specific information about L1 proficiency, comparisons between this study and the work by Proctor et al. are speculative at best.

While the results of this study are intriguing, it is important to note significant limitations. First, the manipulated portions of the experimental texts (approximately 100 words in each of the 200-word texts) may not have been long enough to allow for the detection the effects of syntactic and lexical complexity across all four topics. Additionally, the fact that ELL status was dichotomously classified (ELL or non-ELL) could limit our ability to explore the effect of first language (L1) expertise on performance. A multitude of studies highlight the significant effects of L1 proficiency on L2 acquisition and comprehension (Jimenez, Garcia, & Pearson, 1996; Proctor et al., 2005). The hypothesis that students' command over their first language influences their ability to comprehend both syntactically and lexically complex features of texts was not considered in the present study. Further research is needed to determine possible effects of varying gradations in L1 proficiency on L2 text difficulty.

The findings within the present study have left questions regarding text accessibility unanswered. Does syntactic complexity have absolutely no effect on comprehension, or is there some gradation of difference that was not captured within the design of our texts? Does prior knowledge, as defined by conceptual familiarity, trump lexical complexity, as indexed by frequency, in determining comprehension? If so, how much familiarity is necessary to overcome difficult vocabulary? Finally, do EL learners face the same difficulties as native English speakers in terms of text accessibility, even when considering the effect of gradations in L1 proficiency? We hope that future studies will shed further light on these important questions. At the same time, our failure to elicit a syntactic complexity affect might give us pause, when we design curriculum, of being too rigid about keeping sentence length to an absolute minimum. Further, the lexical complexity effect, which seemed to be most powerful in situations in which students could not rely on prior knowledge from everyday experiences, merits attention for all students who struggle with unfamiliar content when reading in disciplinary settings.

• • •

Diana J. Arya is currently a research fellow working with the Norwegian Center for Science Education at the University of Oslo, Norway. The central focus of her research is on the quality of school science texts and on methods for improving this quality. Arya is a recent graduate from Berkeley (PhD) with additional degrees from Illinois (MA) and Michigan (MA). Before joining the academic research community, Arya was a reading specialist and middle school science
teacher in Seattle, WA.

Elfrieda H. Hiebert is President and CEO of TextProject, Inc., a not-for-profit aimed at increasing student-reading levels through appropriate texts as well as a research associate at the University of California, Santa Cruz. Her research focuses on text design and instruction around the core and extended vocabularies of written English. She began her educational career as a teacher in California and, after receiving the Ph.D. in Educational Psychology from the University of Wisconsin-Madison, taught at the Universities of Kentucky, Colorado-Boulder, Michigan, and California, Berkeley.

P. David Pearson is a Professor in the Language, Literacy and Culture program in the Graduate School of Education at the University of California, Berkeley, where he pursues research on assessment, instruction, and curriculum reform in literacy. With colleagues at Lawrence Hall of Science, he is building and validating an integrated science and literacy curriculum for grades K-8. With degrees from Berkeley (BA) and Minnesota (PhD) and professorial stints at Minnesota, Illinois, and Michigan State, he began his career as a 5th grade teacher in Porterville, CA.

References

Anderson, R.C., & Freebody, P. (1981).Vocabulary knowledge. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp.77–117). Newark, DE: International Reading Association.

Armstrong, J.E. & Collier, G.E. (1990). *Science in biology: An introduction*. Prospect Heights, IL: Waveland Press.

Armbruster, B. (1993). Science and reading. *The Reading Teacher*, 46(4), 346-347.

Bailey, A.L. (2007). Introduction: Teaching and assessing students learning English in school. In A. L. Bailey (Ed.). *Language Demands of Students Learning English in School: Putting academic language to the test.* New Haven CT: Yale University Press.

Beck, I.L., & McKeown, M. (1991). Conditions of vocabulary acquisition. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol., II, pp. 789-814). White Plains, NY: Longman.

Bowey, J.A. (1986). Syntactic awareness in relation to reading skill and ongoing reading comprehension monitoring. *Journal of Experimental Child Psychology,* 41, 282-299.

California Department of Education (2007). *The California Standardized Testing and Reporting (STAR) Program*. Retrieved January 30, 2007 from http://star.cde.ca.gov/star2006/

California State Board of Education (April 17, 2006). *Criteria for evaluating instructional materials (Reading/Language Arts).* Retrieved January 30, 2007 from http://www.cde.ca.gov/ci/rl/im/

Carey, S. (1985). Are children fundamentally different thinkers than adults? In S. Chipman, J. Segal & R. Glaser (Eds.), *Thinking and learning skills* (pp. 436-517). Hillsdale, NJ: Lawrence Erlbaum.

Cervetti, G. and Barber, J. (2009). Bringing back books: Using text to supplement hands-on investigations for scientific inquiry. *Science and Children*, *47*(3), 20-23.

CCSS (2010). Common Core State Standards. Retrieved September 14, 2011, from http://www.corestandards.org/.

Cunningham, A. & Stanovich, K.E. (1998). What reading does for the mind. *American Educator,* 22(1& 2), 8-15.

Donovan, C.A., & Smolkin, L.B (2001). Genre and other factors influencing teachers' book selections for science instruction. *Reading Research Quarterly*, 36 (4), 412-440.

Droop, M. & Verhoeven, L. (1998). Background knowledge, linguistic complexity and second language reading comprehension. *Journal of Literacy Research,* 30(2), 253-271.

Duke, N. (2000). 3.6 minutes per day: The scarcity of informational texts in the first grade. *Reading Research Quarterly,* 35*,* 202-224.

Duke, N. K., & Bennett-Armistead, V. S. (2003). *Reading and writing informational text in the primary grades: Research-based practices.* New York: Scholastic.

Elley, W. (1996). Using book floods to raise literacy levels in developing countries. In V. Greaney (Ed.), *Promoting reading in developing countries: Views on making reading materials accessible to increase literacy levels* (pp. 148-163). Newark, DE: IRA.

Flesch, R. (1979). *How to write plain English*. New York, NY: Harper and Row.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221–23.

Goldman, S.R., & Bisanz, G. L. (2002). Toward a functional analysis of scientific genres: Implications for understanding and learning processes. In J. Ortero, J.A. Leon, & A.C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 19-50). New Jersey: LEA.

Gopnik, A. (1996). The scientist as child. *Philosophy of science*, 63(4), 485-514.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly*, 25(3), 375-406.

Gutierrez, K. , & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, 32(5), 19 - 25.

Guthrie, J. T., Anderson, E., Alao, S., & Rinehart, J. (1999). Influences of Concept- Oriented Reading Instruction on strategy use and conceptual learning from text. *Elementary School Journal, 99(4*), 343-366.

Guthrie, J. T., McRae, A. C., & Klauda, S. L. (2007). Contributions of Concept-Oriented Reading Instruction to knowledge about interventions for motivations in reading. *Educational Psychologist, 42,* 237-250.

Hayes, D. P., & Ahrens, M. (1988). Vocabulary simplification for children: A special case of 'motherese.' *Journal of Child Language,* 15*,* 395-410.

Jimenez, R.T., Garcia, G.E., & Pearson, P.D. (1996). The reading strategies of bilingual; Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly,* 31, 90-112.

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

Klare, G.R. (1984). Readability. In P. D. Pearson (Ed.), *Handbook of Reading Research* (Vol. 1, pp. 681-744). New York: Longman.

Lee, O., & Luykx, A. (2005). Dilemmas in scaling up innovations in science instruction with nonmainstream elementary students. *American Educational Research Journal,* 42(5), 411–438.

Lennon, C. & Burdick, H. (2004). The lexile framework as an approach for reading measurement and success. MetaMetrics.

Leslie, L., & Caldwell, J. (2000). *Qualitative Reading Inventory-III*. New York: Longman.

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., 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 & Instruction,* 14*,* 1-43.

Nagy, W.E., & Scott, J.A.(2000).Vocabulary processes. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research* (Vol. III, pp. 269–284). Mahwah, NJ: LEA.

Nation, K., & Snowling, M.J. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. *Applied Psycholinguistics,* 21, 229–241

National Institute of Child Health and Human Development (NICHD) (2000). *Report of the National Reading Panel. 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). Washington, DC: U.S. Government Printing Office.

National Research Council. (2001). *Classroom assessment and national science education standards*. Washington D.C.: National Academy Press.

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87*(2), 224-240.

Ozuru, Y., Dempsey, K., Sayroo J., & McNamara, D.S. (2005). Effect of text cohesion on comprehension of biology texts. Psychology Department, University of Memphis.

Palincsar, A. S. & Magnusson, S. J. (2001). The interplay of first-hand and text-based investigations to model and support the development of scientific knowledge and reasoning. In S. Carver & D. Klahr (Eds.), *Cognition and instruction: 25 years of progress* (pp.151-194). Mahwah, NJ: Lawrence Erlbaum.

Pearson, P.D. (2009). The roots of reading comprehension. In S.E. Israel & G.G. Duffy (Eds.), Handbook of research on reading comprehension (pp. 3–31). New York: Routledge.

Pearson, P.D., & Camperell, K. (1981). Comprehension of text structures. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 448-468). Newark, DE: International Reading Association.

Proctor, C.P., August, D., Carlo, M., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology,* 97(2), 246-256.

Qian, D.D. (2002) investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective. *Language Learning,* 52, 513-536.

RAND Reading Study Group. (2002). *Reading for understanding: Towards an R&D program in reading comprehension.*

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models*. Thousand Oaks, CA: Sage Publications.

Rawson, K.A. (2004). Exploring automaticity in text processing: syntactic ambiguity as a test case. *Cognitive Psychology*, 49(4), 333-69.

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 four. *Journal of Research in Science Teaching*, 29, 545-554.

Romance, N.R., & Vitale, M.R. (2006). Science IDEAS: Making the case for integrating reading and writing in elementary science as a key element in school reform. In R. Douglas,M. P. Klentschy, K. Worth and W. Binder (Eds.), *Linking science and literacy in the K–8 classroom* (pp. 391-405). Arlington, VA: National Science Teachers Association (NSTA) Press.

Rutherford, J.F. (1991). Vital connections: Children, books, and science. In S.Jagusch & W. Saul (Ed.), *Vital connections* (pp. 21-30). Portsmouth, NH: Heinemann.

Schleppegrell, M.J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: LEA.

Shaw, J. M. (1997). Threats to the validity of science performance assessments for English language learners. *Journal of Research in Science Teaching,* 34(7), 721-743.

Shymansky, J.A., Yore, L.D., & Good, R. (1991). Elementary school teacher's beliefs about and perceptions of elementary school science, science reading, science textbooks, and supportive instructional factors. *Journal of Research in Science Teaching,* 28(5), 437-454.

Smagorinsky, P. (2001). If meaning is constructed, what is it made from? Toward a cultural theory of reading. *Review of Educational Research*, 71(1), 133-169.

Snow, C.E. (2010). Academic language and the challenge of reading for learning about science. *Science*, 328, 450-452.

Snow, C. E., & Sweet, A.P. (2003) Reading for Comprehension. In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (1-11). New York: Guilford Press.

Snowling, M., & Nation, K. (1997). Language, phonology and learning to read. In C. Hulme & M. Snowling (Eds.). *Dyslexia: Biology, cognition and intervention*. San Diego, CA: Singular Publishing Group.

Stahl, S.A. (1999). *Vocabulary development*. Cambridge, MA: Brookline.

Stanovich, K. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: Guilford.

U.S. Census Bureau, (2000). http://www.census.gov/main/www/citation.html.

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). NY: Guilford Press.

Willows, D. M., & Ryan E. B. (1986). The development of grammatical sensitivity and its relationship to early reading achievement. *Reading Research Quarterly,* 21, 253–266.

Wilson, D, & Sperber, D. (1987). An outline of relevance theory. *Notes on Linguistics,* 39, 5-24.

Zeno, S.M., Ivens, S.H., Millard, R.T., & Duvvuri, R. (1995). *The Educator's Word Frequency Guide*. Brewster: Touchstone Applied Science Associates, Inc.

International Electronic Journal of Elementary Education, 2011, 4(1), 127-152.

Structure strategy interventions: Increasing reading comprehension of expository text

Bonnie J. F. MEYER ∗
The Pennsylvania State University, United States

Melissa N. RAY
The Pennsylvania State University, United States

Abstract

In this review of the literature we examine empirical studies designed to teach the structure strategy to increase reading comprehension of expository texts. First, we review the research that has served as a foundation for many of the studies examining the effects of text structure instruction. Text structures generally can be grouped into six categories: comparison, problem-and-solution, causation, sequence, collection, and description. Next, we provide a historical look at research of structure strategy interventions. Strategy interventions employ modeling, practice, and feedback to teach students how to use text structure strategically and eventually automatically. Finally, we review recent text structure interventions for elementary school students. We present similarities and differences among these studies and applications for instruction. Our review of intervention research suggests that direct instruction, modeling, scaffolding, elaborated feedback, and adaptation of instruction to student performance are keys in teaching students to strategically use knowledge about text structure.

Keywords: Text structure, structure strategy intervention

Introduction

Reading for understanding is vital for readers of all ages. The 2009 National Assessment of Educational Progress (U. S. Department of Education National Institute for Education Statistics, 2010) reported that 33% of 4th-grade students examined could not read at the basic level required to understand what they read. Comprehension of expository text is critical for academic success in school (National Educational Goals Panel, 1999). Despite its importance, in comparison to narrative texts, students receive less exposure to expository

∗ Bonnie J. F. Meyer, Department of Educational and Counseling Psychology and Special Education, 204 Cedar Building, Penn State, University Park, PA 16802, United States. Phone: 814-863-7501; FAX: 814-863-1002. E-mail: bjm8@psu.edu

ISSN:1307-9298

www.iejee.com

texts in early elementary school (e.g., Duke, 2000). This lack of exposure may place readers at a disadvantage because beginning in 4th-grade students increasingly are expected to learn from expository texts in language arts, science, and social studies (Guthrie & Davis, 2003).

Reading comprehension involves actively constructing new understandings by building relationships among the parts of text and between the text and one's pre-existing knowledge. Good readers build coherent mental representations of what they read by understanding different text structures, generating inferences, monitoring their understanding, and using multiple strategies to construct meaning. Use of text structure to understand how the important ideas of a text are inter-related increases readers' meaning making. Readers who use text structure can mentally examine how ideas in text are inter-related through the use of such relationships as sequence, comparison, causation, or problem and solution. These readers also may use external aids that show the top-level structure of a text to reduce memory demands. These aids include templates, text structure patterns, graphics, matrices, outlines, knowledge maps, or tree structures (e.g., Meyer, Young, & Bartlett, 1989).

Over the last 40 years, Meyer (e.g., 1971, 1975) and her colleagues (e.g., Meyer & Rice, 1982; Meyer et al., 1989; Meyer & Poon, 2001; Meyer et al., 2010) have studied text structures and readers' abilities to use them (see Table 1). Readers who use a "structure strategy" seek to identify and use the author's organization to organize their own understanding (Meyer, Brandt, and Bluth, 1980). The structure strategy facilitates comprehension by helping the reader to organize concepts based on the explicit or implied relationships that are communicated by the text. The strategy promotes comprehension compatible with van den Broek's coherence-based processes in his simulation model of comprehension (van den Broek, Rapp, & Kendeou, 2005). The causal connections of his model focus on the important cause-and-effect relationships that make up the logical structure of narrative text just like text structures build on each other to establish the logical structure for nonfiction text. Text structures not only describe the text itself, but also characterize readers' cognitive coherence representations (e.g., Meyer & Freedle, 1979, 1984; Sanders & Noordman, 2000). Good readers use their knowledge of text structures to build coherent memory representations (Meyer et al., 1980). Signaling words (see Table 1) can cue text structures and assist readers toward building coherent text representations. The key role of signaling words is in selection and encoding, particularly if readers have learned the structure strategy (e.g., Meyer & Poon, 2001).

The power of teaching students the structure strategy is that it enables them to a) follow the logical structure of text to understand how an author organized and emphasized ideas; b) use processes parallel to these structures to increase their own learning and thinking (e.g., comparing, finding causal relationships, looking for solutions to block causes of problems); and c) use these text structures to organize their own writing, such as written summaries, recalls, and essays.

In this paper we review empirical studies of interventions designed to teach the structure strategy in order to improve reading comprehension. First, we briefly discuss the basic research that served as a foundation for instruction about text structures. Next, we provide a historical look at structure strategy interventions. In this historical overview, we examine both the history and advancements in studies of structure strategy instruction. Finally, we review recent text structure interventions for elementary school students. For each intervention study included in this review, we examine the strategy instruction in terms of the number and types of text structures and/or signaling words taught, the methods of instruction, the rigor of the intervention design, and relevant findings.

Table 1. *Text Structures with Signaling*

COMPARISON (compare/contrast) – relates ideas by differences and/or similarities; complexity can be increased by the number and detail of issues compared. The main ideas are organized to provide a comparison, contrast, or alternative view (e.g., political speech).

instead, but, however, alternatively, whereas, on the other hand, while, compare, in comparison, in contrast, in opposition, not everyone, all but, have in common, similarities, share, resemble, the same as, just as, more than, longer than, less than, act like, look like, despite, although, difference, differentiate, different…

PROBLEM-and-SOLUTION – relates responding ideas; complexity can be increased by the identification of causes of the problems and ways to reduce them. The main ideas are organized in two parts: the problem (or question) part and the solution (or answer) part, which responds to the problem part (e.g., popular science articles, medical information).

Problem: problem, trouble, difficulty, hazard, need to prevent, threat, danger, puzzle, can hurt, not good, bad…

Solution: to satisfy the problem, ways to reduce the problem, so solve these problems, protection from the problem, solution, in response, recommend, suggest, reply…

CAUSE-and-EFFECT (causation) – relates ideas casually; complexity can be increased by embedded cause and effect paths and causal chains and reduced by similarity to familiar narratives. The main ideas are organized into cause and effect parts (e.g., directions, explanations, economic or science texts)

cause, led to, bring about, originate, produce, make possible, owing to, by means of, accomplish by, since, due to, because, in order to, reasons, why, if/then, on account of, in explanation, effect, affects, so, as a result, consequence, thus, therefore, accordingly, for the purpose of…

SEQUENCE (time ordered collection of events, ideas) – relates ideas via time. The main ideas are the steps or history presented (e.g., recipe steps, history books, biographies)

later, afterwards, after, then, subsequently, as time passed, following, continuing on, to end, finally, year(s) ago, at the start of, first, second, third, 1, 2, 3…, next, primarily, early, before, to begin with, more recently, again, finally, the former, the latter, not long after, soon, now, today, after a short while, meanwhile, steps, stages, time line, history, sequence, development, look for a series of dates in histories…

COLLECTION (listing, enumeration) – relates ideas simply by grouping them together; sometime the grouping is made explicit with enumeration. The main idea is the grouped list (e.g., "to do list," botany). Collection can be used with any of the other structures; sequence is a subtype of collection. For example, groups of solutions or causes are often presented.

and, in addition, also, include, moreover, besides, first, second, third, etc., subsequent, furthermore, at the same time, another , and so forth…

DESCRIPTION (generalization, settings) – relates ideas by elaboration of attributes, specifics, or setting information. The main ideas is that aspects of a topic are presented (e.g., newspaper article)

attributes of, characteristics are, for example, for instance, in describing, marks of, namely, properties of, qualities are, specifically, such as, that is…

Basic Text Structure Research

Historical antecedents of many structure strategy interventions were basic laboratory studies about text structure and its effects on learning and memory (Meyer & McConkie, 1973; Meyer, 1975; Meyer & Freedle, 1984; Meyer & Rice, 1982). Meyer (1971) identified the hierarchical, logical structure of two articles from *Scientific American* magazine and examined the relationship between the structure and what college students remembered from the text. Effects of this logical structure were seen in a) the kinds of idea units that were remembered (more high level information than low level information; this effect came to be known as the "levels effect"); b) the stability of the idea units in consecutive recalls (stability related to the logical structure [.55], rather than rated importance [-.25]); and c) the tendency for clustering on this basis (if a particular idea unit was recalled, then the idea directly above it the logical structure was recalled 70% of the time; overall recall was 23% of ideas from a passage) (Meyer, 1971). Logical structure accounted for much of the variance that might ordinarily be attributed to other variables, such as serial position effects and rate importance. These findings suggested that the logical structure of a passage is related to certain aspects of the cognitive structure that the participants constructed.

Next, Meyer (1974, 1975) combined the logical structure approach of the initial study with work by Joseph Grimes (1975) in linguistics. This provided methods for studying naturally occurring text and ways to control aspects of text structure and signaling for future experiments (see simplification of the approach in Meyer, 1985). Other text analysis methods at the time included work by Crothers (1972) and Frederiksen (1972, 1975). An important manipulation by Meyer (1975) was the embedding of a causation (cause-and-effect) paragraph in the same serial position in two texts. In one text the paragraph was located at the top third of the hierarchical, logical text structure. In the other text, the paragraph was located at the bottom third of the text structure (see Figure 1 for example of hierarchical, logical structure of a text). This manipulation was repeated with a text with the same structure but different content. Also, versions of the set of four texts were prepared with and without signaling words that explicitly signaled the text structures so that patterns in recall of ideas could be compared as they varied in content, structure, or level in the logical structure. Meyer found that the type and structure of relationships among concepts in text dramatically influenced comprehension when they occur at the top third of the structure. However, when the same pattern of relationships occurred low in the structure, they affected comprehension minimally (Meyer, 1975). This was an important finding because it encouraged future text structure research to focus on the main ideas organized within a text rather than the details.

Due to this focus on top-level of text structures, Meyer (1975) hypothesized that more organized text structures (causation, comparison, and problem-and-solution) would have greater mnemonic hooks for learning and memory than the structure of description (i.e., a collection of descriptions about a topic). Meyer and Freedle's (1979, 1984) data supported this hypothesis, showing that for college learners, comparison and causation structures had greater benefits for recall in comparison to description. Later work showed that the memory benefit for the more organized structure of comparison over a collection of descriptions held for adults who could use the structure strategy and who had high average and above vocabulary skills (Meyer et al., 1989). However, for young, middle-aged, and older adults with average vocabularies and no training in the structure strategy, the collection of descriptions yielded better recall than the comparison structure (Vincent, 1985). This effect of text structure has also been found with college level English as second language (ESL) readers (Carrell, 1984).

By 1976, Kintsch's hierarchical, text base analysis was becoming popular. The structure developed by Kintsch and colleagues (Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975) was based on argument repetition (i.e., repetition of words from the text), while Meyer's hierarchy was based on the semantic relations among ideas represented by text structures (see Figure 1 and Table 2). Dunn (1978) stated,

> "Unlike Kintsch's system, Meyer's system not only produces a hierarchical arrangement, but also states explicit inter-propositional relationships (including inter-paragraph relations among the items in text) to a much greater degree than do either Kintsch and his associates (e.g., Turner & Green, 1977) or Frederiksen…" (p. 10)

In 1976 Meyer directed her research efforts to the strengths of her approach; that is, representations of the top-level structures in text. Several research efforts were launched to investigate the differences in use of the structure strategy among students with varying reading and vocabulary proficiencies and ways to increase use of the structure strategy by changing the text (e.g., signaling), the task (e.g., repeated readings of text with different content but the same text structure), or the reader (e.g., various levels of instruction).

Meyer, Brandt, and Bluth (1978, 1980) studied whether good and poor 9th-grade readers organized their recall with the same text structure as that used by the author. Good and poor readers were identified by scores on a standardized reading comprehension test and corroborated by teacher appraisals. Two basic strategies were expected from the 9th graders: the structure strategy or the default/list strategy. While the structure strategy involves systematic processing based on text structure, the default strategy is not systematic. The reader using the default strategy lacks focus and simply tries to remember some ideas from the text. The recall produced by such a reader is a list-like collection of descriptions about the topic with little to no attempt to interrelate the ideas. Meyer et al. (1980) also examined the effects of signaling words for problem-and-solution and comparison text structures (see underlined words in Table 2 for the problem-and-solution text).

Table 2. *The Supertanker Text.*

(*Note: CAPITALIZED = author's message or main ideas; lowercase = major details; italics = minor details; underlined = signaling. In actual studies regular font was used for the text.)*

A PROBLEM OF VITAL CONCERN IS PREVENTION OF OIL SPILLS FROM SUPERTANKERS. A typical supertanker carries a half-million tons of oil and is the size of five football fields. A wrecked supertanker spills oil into the ocean; this oil kills animals, birds, and microscopic plant life. *For example, when a tanker* crashed *off the coast of England, more than 200,000 dead seabirds* washed *ashore*. Oil spills also kill microscopic plant life *which provide food for sea life and produces 70 percent of the world's oxygen supply*. Most wrecks RESULT FROM THE LACK of power and steering equipment to handle emergencies, *such as storms*. *Supertankers have only one boiler to provide power and one propeller to steer the ship.*

THE SOLUTION TO THE PROBLEM IS NOT TO IMMEDIATELY HALT THE USE OF TANKERS ON THE OCEAN since about 80 percent of the world's oil supply is carried by supertankers. INSTEAD, THE SOLUTION LIES IN THE TRAINING OF OFFICERS OF SUPERTANKERS, BETTER BUILDING OF TANKERS, AND INSTALLING GROUND CONTROL STATIONS TO GUIDE TANKERS NEAR SHORE. First, officers of the supertankers must get top training in how to run and maneuver their ships. Second, tankers should be BUILT with several propellers *for extra control* and backup boilers *for emergency power*. Third, GROUND CONTROL STATIONS SHOULD BE INSTALLED at places where supertankers come close to shore. These stations would act like airplane control towers, guiding *tankers along busy shipping lanes and through dangerous channels.*

Source: Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of the top-level structure in text: Key for reading comprehension of ninth-grade students. Reading Research Quarterly, 16, 72–103. Copyright @ 1980 by the International Reading Association (www.reading.org).

For a subgroup of readers identified as comprehension "underachievers," signaling was expected to be particularly helpful. Students in this subgroup had better vocabulary (stanine score at least 4) than reading comprehension skills (one stanine below a student's vocabulary score). While "underachievers" had vocabulary test scores closer to the good readers, reading comprehension scores of the underachievers were closer to the poor readers. They were identified as readers who could use the structure strategy, but would not without the explicit prodding provided by the signaling words.

Meyer et al. (1980) found that reading ability was associated with the number of ideas recalled, as well as the organization of recalls. Readers classified as good readers recalled more ideas, and more frequently used the authors' structure to organize their recalls. The researchers also found that text signaling influenced underachievers text processing. When reading text with signaling, as shown in Table 2 and Figure 1, these readers switched to use of the structure strategy instead of the default/list strategy. Signaling did not affect the use of the structure strategy by good readers or poor readers. Regardless of random assignment to the signaling or no signaling conditions, good readers recalled the top-level structure of the passage and used it to organize their recall, while most poor readers did not (Meyer et al., 1980). This finding pointed to the importance of teaching signaling words as part of instruction in the structure strategy for students with poor reading comprehension skills. In this study, only 48% of the entire sample of 9th-grade students organized their recall with the same structure as the text on at least one of the problem-and-solution and comparison texts (Meyer et al., 1980).

From this text structure research, Meyer and collaborators (Meyer & Freedle, 1979, 1984; Meyer, 1984; Meyer & Rice, 1982) developed a processing model for getting text information into organized schemata for storage in memory based on text structures. First, readers determine whether they are interested in communicating with the writer of the text and following the writer's thesis and rationale. If not, readers should use a different reading strategy rather than the structure strategy. Next, readers select the structure strategy. If readers cannot use this strategy, the default/list strategy by default would be used. Figure 2 depicts the processing steps in using the structure strategy. The end point in the model is a reader using the identified working schema, corresponding to a text structure, as an organizing framework to differentially select, encode, and organize ideas from the text into a long-term memory representation.

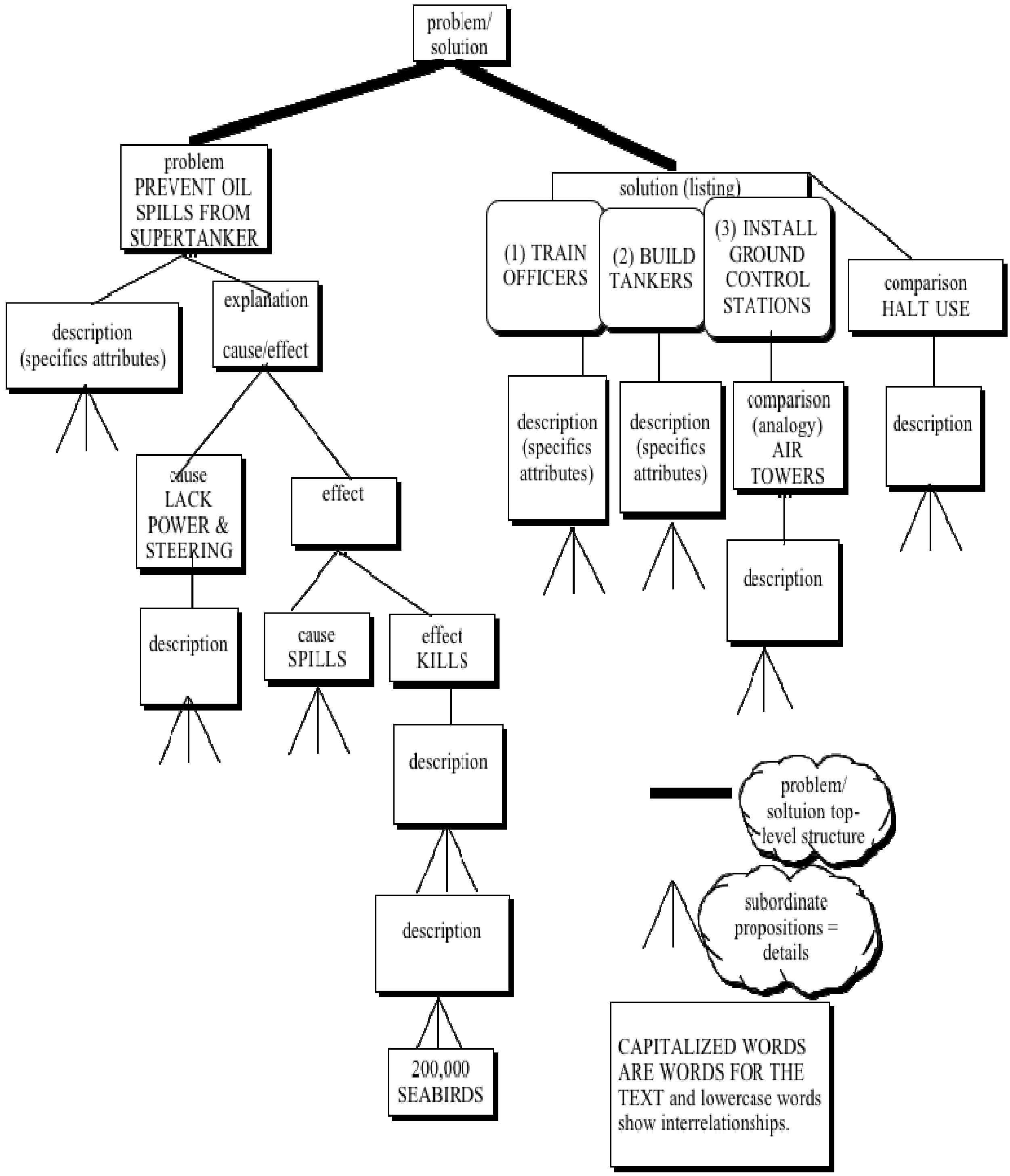
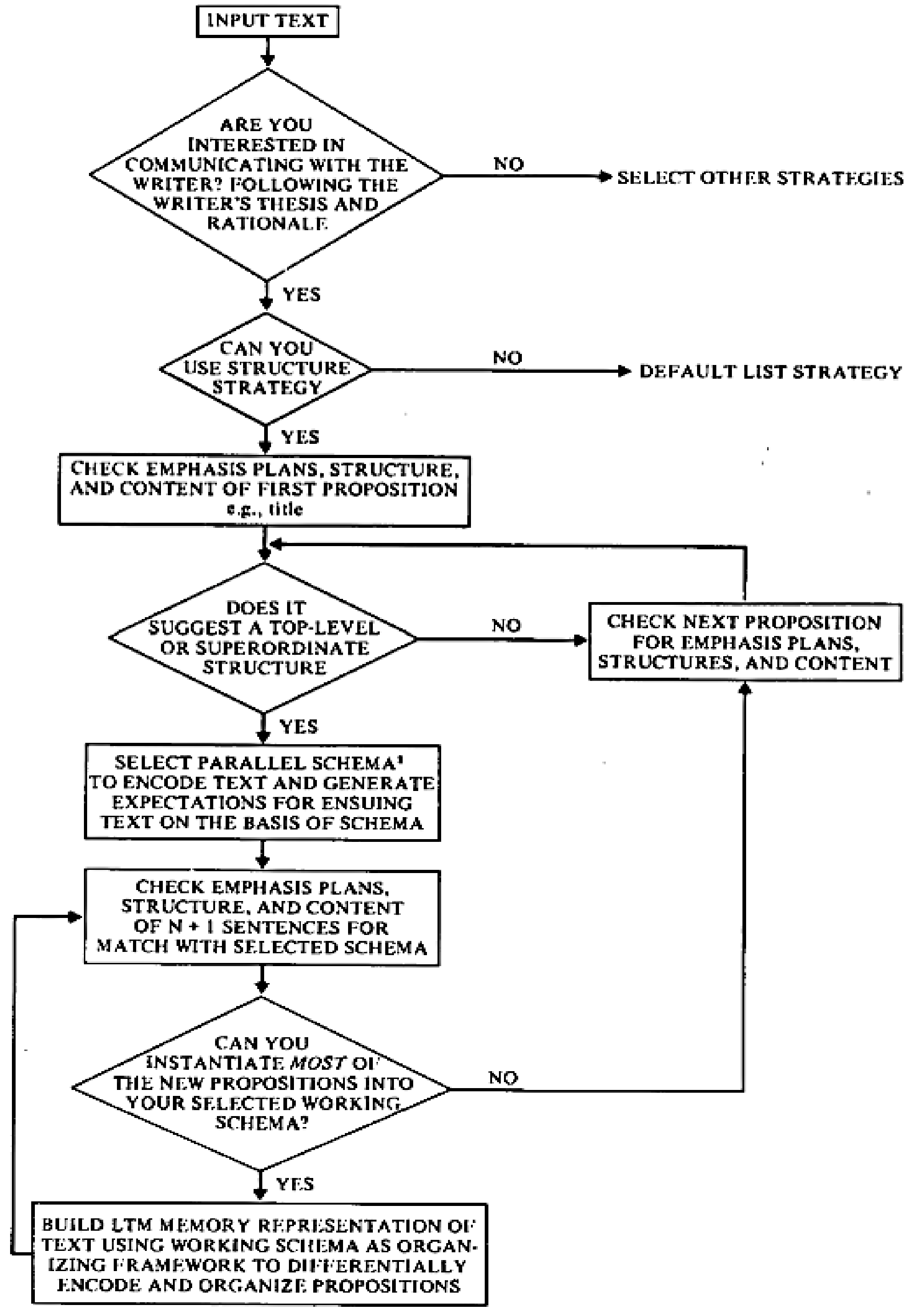


Figure 1. *Top-Level Structure of the Supertanker Text. From Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of the top-level structure in text: Key for reading comprehension of ninth-grade students. Reading Research Quarterly, 16, 72-103. Copyright @ 1980 by the International Reading Association (www.reading.org).*



[1] Type of schema selected here influences processes of selection and buffer rehearsal.

Figure 2. Model for getting text information into organized schemata for storage in memory. From Meyer, B. J. F. (1984). *Text dimensions and cognitive processing*. In H. Mandl, N. Stein & T. Trabasso (Eds.), *Learning and comprehension of text*. Hillsdale, NJ: Lawrence Erlbaum Associates. Permission: *Learning and comprehension of text* by Mandl, Heinz. Copyright 1984 Reproduced with permission of TAYLOR & FRANCIS GROUP LLC - BOOKS in the format Journal via Copyright Clearance Center.

Empirical studies of text structure and its effect on text processing and comprehension had several important implications for future intervention research. First, they provided evidence that text structure indeed exerted an influence on readers' mental representations of texts. Second, they suggested that readers vary in their ability to use the structure strategy, and this variability may be related to their overall reading ability and vocabulary. Finally, the findings from Meyer et al. (1978, 1980) suggested that many readers may be lacking in their knowledge of text structures. Overall, this basic research provided evidence that many readers would likely benefit from explicit instruction in the use of text structure and encouraged studies which investigated methods of teaching readers to use them.

Early text structure interventions (1978-1990)

In this section we discuss the early history of text structure interventions. These began in the 1970s with several doctoral dissertations or projects under Meyer's mentoring, which examined instruction designed to increase middle school to junior college students' use of the structure strategy (e.g. Brandt, 1978; Jessen, 1981; Meyer, Bartlett, Woods, 1978). Bartlett's dissertation (1978) was the first study to provide extended multiple sessions of explicit structure strategy instruction; he taught 9th-grade students Meyer's expository discourse types of problem/solution, comparison, causation, and collection of descriptions. All four structures were taught each day for five days of 1-hr instruction. Bartlett's approach asked students to find the main idea and then determine the text structure that organized the main idea. Emphasis was not placed on signaling words or patterns/templates to write main ideas or recalls. Explicit instruction of 6-steps to follow before, during, and after reading to use the top-level structure strategy were modeled and practiced. Texts used as examples for instruction and practice increased in complexity across the five sessions.

Structure strategy instruction increased students' ability to identify and use the text's top-level structure and nearly doubled the amount of information remembered over students in the control condition who received the same texts, but with instruction about punctuation (Bartlett, 1978). Instruction effects appeared durable over an extended period for readers scoring above the 19th percentile on a standardized vocabulary test (Bartlett, 1978). Teachers reported performance advantages across the curriculum for students who had received training with the structure strategy rather than the punctuation instruction.

Armbruster also worked with Meyer and her approach to text structure in the late 1970s and 80s. Subsequently, Armbruster, Anderson, and Ostertag (1987) taught the problem-and-solution structure to 5th-grade students in 11 days of direct instruction with social studies materials. One of two classrooms of students was randomly assigned to structure strategy training with the problem-and-solution structure and the other classroom to traditional instruction. Students in both groups read 13 100 to 500-word texts taken from 4th- and 5th-grade textbooks. The structure strategy group received problem-and-solution frames accompanied by blank lines for writing passage summaries. The frames were three boxes with an arrow from the problem box (something bad) pointing towards the action box (what people do attempting to solve the problem) and the results box (outcome of the action). A second arrow went from the action box to the results box. Instructors modeled how to identify the text structure and how to use the problem-and-solution structure to write a summary, following principles of direct instruction (e.g., Rosenshine & Stevens, 1986). Guidelines were provided for writing a summary of a problem-and-solution passage. These guidelines prompted students to write the problem, the solution, and the result of this solution, but did not emphasize problem and solution signaling words. Students in the structure strategy class wrote 50% more main ideas on an essay exam and included more main ideas in written summaries of a text than the traditional class, but the classes did not differ on a short-answer fact test (Armbruster et al., 1987). Although the study's experimental design was limited in rigor, it provided instruction with a 5th-grade classroom setting with authentic texts.

In addition to these interventions, several early interventions investigated structure strategy interventions with adult learners. These studies covered a diverse group of learners including English language learners, college students, and adults learning within a community setting. Carrell (1985) generally followed Bartlett's (1978) training materials with the same four text structures outlined by Meyer (e.g., Meyer & Freedle, 1984). Unlike Bartlett, Carrell instructed high-intermediate ESL students from various language backgrounds enrolled in an intensive English program. One section of 14 students received structure strategy instruction, while the other section of 11 students read the same reading materials, but worked with linguistic operations, such as sentence combining, cohesion, and vocabulary. Similar to Bartlett, Carrell found that the structure strategy group showed substantially and significantly higher performance on measures of reading comprehension than the control group after training as well as three weeks later. Carrell's study was the first to show that direct instruction with the structure strategy instruction increased reading comprehension of ESL students.

In their intervention, Cook and Mayer (1988) built on the work of Meyer (Meyer, 1975; Meyer et al., 1980) by adding a number of important components to an instructional program which taught college students to use text structure to comprehend science texts. These components included a) a sorting task to measure text structure awareness, b) the application of text structure to passages taken from students' chemistry textbooks, and c) specification of descriptive text structures that occur in science textbooks. Cook and Mayer conducted two studies about text structure. They added a novel component of a sorting task in which students sorted 20 texts taken from high school science textbooks according to five expository text structures. In the first study, half of a group of 32 undergraduate university students received a 5-page instructional pamphlet about five text structures used in science textbooks: generalization, enumeration, sequence, classification, and compare/contrast. Cook and Mayer presented a couple of signaling words per structure for the classification (i.e., "there are two types"), comparison (i.e., "in contrast to"), and sequence ("and then") structures. The text structures classified by Cook and Mayer, particularly, generalization and classification, differ somewhat from Meyer's classification, but could be subsumed as subtypes of description. The students in the two groups worked at their own pace sorting texts into similar groups based on structure. The training group sorted the 4 texts per 5 text types correctly 79% of the time, while the no training group sorted them correctly 61% of the time.

The second study conducted by Cook and Mayer (1988) involved training junior college students and provided instruction about three of the five text structures. Students filled out three worksheets for the three trained text structures (generalization, enumeration, and sequence) using nine passages taken from their chemistry textbook. For example, the sequence worksheet had three steps: step 1 – identify the passage topic; step 2 – name each step in the sequence and outline the details of each; step 3 – say what varies from one step in the sequence to the next. Another section of the chemistry class received no training and served as the control group. Trained students increased their recall of the most important information and ability to answer application questions, but did not increase memory for facts. The results demonstrated the value of the structure strategy using an outlining format with description, listing, and sequence structures in the context of a science class.

The research about knowledge-map (k-maps) by Dansereau and colleagues (e.g., Dansereau et al., 1979; Holley, Dansereau, McDonald, Garland, & Collins, 1979) has similarities to the structure strategy. We will only briefly mention this large research literature (see O'Donnell, Dansereau, & Hall, 2002). To make a k-map from an existing text the creator of the k-map needs to identify and use text structures. Four of the links in k-maps are "part of," "type of/example of," "characteristic of," and "evidence for;" these are subtypes of the description text structure. The other two links are "leads to" and "analogous to;" they correspond to causation and comparison, respectively. Holley et al. (1979) provided college students with 5.5 hours of training over four sessions about these links applied to sentences, texts, and their own textbooks. Students with the k-map training performed better than a control group of students on multiple measures of reading comprehension associated with understanding main ideas. Geva's (1983) flowcharting of expository text is a combination of representations of text by Meyer et al. (1980) and Holley et al. (1979). Geva trained community college students to represent text in node-relation flowcharts. Relations were represented by different types of lines in a flowchart matching a key with relations of elaboration, cause-effect, process, example, and detail as well as topic and conclusion. Findings indicated that training with this approach led less skilled readers to more carefully read expository text and as result increase their reading comprehension. Mayer (1999) considered k-maps as an effective type of structure strategy along with Cook and Mayer's (1988) intervention and Meyer et al.'s (1980) hierarchical text structures displayed in Figure 1.

In addition to academic settings, early research examined the efficacy of structure strategies in non-academic settings. Meyer, Young, and Bartlett (1989) utilized some of the texts, feedback materials, and formats from Bartlett's training program in an instructional program designed for young (18 to 32 years old) and older (65 years and older) adults and primarily used everyday reading materials. Instruction consisted of 7.5 hours of instruction (5 sessions) spread over two weeks Instructors and peers (old and young adults) modeled the use of templates unique to each text structure in the composition of written main ideas and recall protocol. Participants were instructed to "choose it (overall/top-level text structure), use it, or lose it." The design of this study was the strongest of extant structure strategy intervention studies at the time. Participants were randomly assigned to three groups: structure strategy instruction, a contact control that practiced reading and recalling the same texts that were used in the structure strategy instruction, but with no instruction about text structure, and a no contact, wait-list control group. Alternate equivalent forms of measures were counterbalanced over a pretest, immediate post-test, and two-week delayed post-test. The instruction changed from Bartlett's earlier approach by incorporating use of signaling words to help identify text structures in simple advertisements to more complex materials (e.g., magazine articles).

Another major change from Bartlett's instruction (1978) was the addition of an initial step in which students were prompted to first search for the top-level structure of the text that could interrelate all the ideas in the text. This top-level structure would then lead to the main idea of the text. The main idea was identified as the ideas interrelated by the top-level structure. This step was added to assist the reader in constructing a coherent representation or situation model.

Some interesting findings regarding elements of effective instruction emerged from the Meyer et al. (1989) study. There were minimal effects of training when students were simply told the definition of different text structures along with some of their signaling words and an example of each text type. More intensive instruction, including modeling how to use text structure strategically for understanding and remembering, appeared to be required to see strong effects of structure strategy training. Often composition or reading textbooks simply mention different text structures; this is not sufficient for increasing reading comprehension. This brief treatment of text structures in such textbooks may relate to barriers in getting teachers to use the structure strategy in elementary and middle school classrooms. Teachers may have encountered a cursory examination of text structure in college classes for composition or reading instruction, but did not see its value for increasing reading comprehension of their students. Thus, they may discount structure strategy interventions as recommended by research or colleagues who have used the strategy. This discounting may be attributed to a lack of depth in understanding the structure strategy as well as a lack of materials needed for modeling and direct instruction about strategically using text structure to increase reading comprehension. Some of the recent structure strategy interventions use intelligent tutors or scripted lessons for teachers, which may overcome some of these obstacles.

Overall, the early work on structure strategy instruction showed its potential for increasing reading comprehension. Positive effects for using the structure strategy were noted from work with elementary school children to retired adults. Most of the research involved 6th graders, 9th graders, high school students, college students (including junior college and ESL students) and adults, rather than early elementary school children. Most instruction programs involved modeling, practice, direct instruction, scaffolding, and multiple instructional sessions of increasing complexity of text materials.

Recent Developments in Structure Strategy Interventions

Text structure instruction across cultures and languages. More recent intervention research has explored the instruction of text structure with linguistically diverse populations. Several studies have explored instruction of expository structure in languages other than English including French (Raymond, 1993), Spanish (Leon & Carretero, 1995), and Dutch (Broer, Aarnouste, Kieviet, & Leeuwe, 2002). One of the first of these was Raymond (1993) who used the instruction and texts from Meyer et al. (1989) translated into French. As with the studies of Bartlett (1978) and Carrell (1985), Raymond randomly assigned two intact classrooms to either structure strategy instruction or a control group that read the same materials as the strategy group, but without instruction about text structures. In Raymond's study the control students worked on answering questions about the texts. The structure strategy training was in French and presented to native English speakers with high-intermediate level skills in French. Participants read articles for the pretest and post-test in French, but recalled them in English. The structure strategy group outperformed the control group on number of ideas recalled on the post-test.

Similarly, Leon and Carretero (1995) conducted an intervention to teach the structure strategy to high school students (ages 14 to 15 years) in Spain, adapting the instruction of Meyer et al. (1989) and a dependent measure from Meyer (1984). They examined the effect of reading comprehension skills (good and poor readers), the text structure intervention in Spanish, signaling, and time of post-testing. There were a number of interesting interactions among these variables. Overall, direct instruction about the structure strategy improved reading comprehension over the control groups who read the same social studies materials, but without instruction in the structure strategy. Additionally, the structure strategy instruction transferred to a text structure not studied in the intervention.

In their study of Dutch 6th graders, Broer, Aarnoutse, Kieviet, and Leeuwe (2002) provided structure strategy instruction for two text structures in a treatment they called the 'making schematics' strategy. The schematics were graphics for the causation structure and the classification structure (similar to Cook and Mayer's [1988] enumeration and a subset of Meyer and Freedle's [1984] collection of descriptions). For example, the schematic for causation was a table headed with a subject row for the topic of a text to be placed. The next row in the table was for signaling words (i.e., "cause," "causes,", "result," and "led to"). The following row was divided into three parts: a column for cause(s), a box for a causal arrow, and a column for result(s). The next row was blank for students to write causes and effects found in a text. The final row was a place to write the main idea organized with causation. Many aspects of the schematic are similar to classroom applications of the structure strategy (e.g., Meyer, Ireland, & Ray, 2011).

Broer et al. (2002) used a non-equivalent control group design with multiple schools, 18 classes, and 354 6th-grade students from middle class homes in The Netherlands. Pairs of classes in a school were randomly assigned to the structure strategy or traditional Dutch reading and answering questions approach. There were 16 instructional lessons taught by the classroom teachers who were trained in the strategy. Students in the schematic/structure strategy condition increased their recognition of text structure, ability to make schematics, ability to formulate and deduce main ideas, and transfer of deducing main ideas from text with different top-level structures than those explicitly taught in the instruction. Effect sizes between the structure strategy and control condition ranged from 0.15 (on a general reading comprehension test) to 0.26 (main idea text) to 0.78 (making schematics four weeks after instruction). The Broer et al. (2002) study was an impressive applied study of structure strategy instruction delivered by classroom teachers at the upper level of elementary schools. All three of these studies showed the usefulness of the structure strategy across cultures.

Recently, Schwartz and colleagues (Mendoza & Schwartz, 2011; Yeh, Schwartz, & Baule, in press) conducted two interventions using the structure strategy via the training materials and testing materials of Meyer et al. (1989) with bilingual college students. One new aspect of this research is the examination of the effects of the structure strategy instruction on eye-movement patterns. Eye-tracing measures examined online processing during learning from texts. They indicate not only what a participant looks at but also how long the person's gaze remains at a particular point and how the eye moves over the text. Yeh, Schwartz, and Baule (in press) reported that after the text structure instruction recall of text information increased and eye-movement patterns changed. After instruction, English for speakers of other languages (ESOL) students made more fixations on the signaling words (e.g., "different" for the comparison text structure) and key areas of the text. The study demonstrated the effectiveness of the structure strategy for ESOL students.

Additionally, Mendoza and Schwartz (2011) taught bilingual college students from the UTEP and Universidad Autónoma de Ciudad Juárez the structure strategy. Results showed that these Spanish speakers also could learn the structure strategy in English. Interestingly, most students were able to transfer their new knowledge about text structures and their English signaling words to parallel structures and signaling in their native (Spanish) language. Texts used as the dependent measures were written in English and Spanish about topics relevant to prenatal screening.

Researchers who have examined the use of the structure strategy in diverse language contexts have included both elementary and college students. Their research adds to the earlier investigations by showing that instruction of the structure strategy aids reading comprehension for bilingual students and students in various cultures.

Structure strategy, signaling, and transfer to everyday learning. In an extension of the earlier work by Meyer et al. (1989), Meyer and Poon (2001, 2004) examined the interaction between structure strategy instruction and signaling in text. Their instructional program, which targeted younger and older adults, provided more instructional emphasis on writing with the use of templates for writing main ideas or recalls with each text structure. These writing patterns have become important components of our recent work with children. For example, for the comparison structure the pattern for writing a main idea is located below. "The main idea was ______ and ______ (2 or more ideas) were compared on _____, ______, _____ (a number of issues)." The pattern for writing a recall from a text with a comparison top-level structure would include a) an introductory sentence with a comparison signaling word contrasting two ideas or political candidates; b) a paragraph or more about the first idea describing the issues for this idea or candidate; c) a transitioning signaling word, such as "In contrast," as a new paragraph is started to describe the second idea or candidate on the same issues. Meyer and Poon also added a sixth instructional session due to needs of older adults (Meyer, Talbot, Stubblefield, & Poon, 1998); this session applied the structure strategy to note-taking, long, unedited magazine articles, medical decision-making, articles from the Internet, and watching an informative video on nutrition. Younger and older adults participated in nine hours of either structure strategy or interest strategy training (reading the same texts as the structure strategy group, but with a motivation strategy), or no training (waiting-list control group). Participants also were randomly assigned to texts with or without signaling for pretests and post-test.

Both training groups reported positive changes in reading, but only the structure strategy group showed increased total recall from a variety of texts (d = 0.64), an informative video (d = 1.47), and information from the medical decision-making task (d = 0.93). The structure strategy intervention affected the organization of recall and was critical for producing readers who could use the structure strategy consistently across a variety of expository texts. The instruction also helped learners to use signals in text more effectively in order to employ the structure strategy across five passages consistently. When compared to the use of signaling without structure instruction, structure strategy training had a larger effect on reading comprehension. However, the relationship between strategy instruction and signaling was shown to be additive; structure strategy instruction plus signaling in texts produced more consistent use of the strategy across five texts. Results also indicated that signaling effects the encoding processes rather than retrieval processes. This was the first study with the structure strategy to show transfer from structure strategy training with texts to multimedia learning and remembering medical information during a simulated decision-making scenario. Moreover, this study, along with Meyer et al. (1989) and Meyer, Talbot, Poon, and Johnson (2001) demonstrated that successful structure strategy instruction could be conducted in a variety of both formal and informal educational contexts.

Structure Strategy Interventions in Elementary Schools

Web-based structure strategy instruction. In their studies of adult learning Meyer and colleagues frequently received feedback from adult participants who voiced their belief that the structure strategy instruction would be of particular use to their grandchildren coping with learning from texts in the school context. The first investigation to follow up on this suggestion (Meyer et al., 2002) involved 5th-grade students tutored by older adults in the structure strategy via the Internet. This invention became the first of many structure strategy interventions targeted for elementary school readers, as part of a renewed interest in investigating structure strategy instruction with younger readers.

Meyer et al. (2002) developed a Web-based delivery of the structure strategy intervention for 5th-grade readers based on the program developed by Meyer et al. (1989). In this intervention five text structures were introduced sequentially rather than all at once because previous work (Meyer, Poon, Theodorou, Talbot, & Brezinski, 2000) had found that adults with slightly reduced working memory resources learned the structure strategy better with lessons introducing one or two text structures per training session versus all five at once. In Meyer et al. (2002) the comparison and problem and solution structures were taught first followed by causation, sequence, and description. These earlier structure were then reviewed in later lessons and integrated implicitly and explicitly with other text structures (cause and effect, sequence, and description) throughout a set of 25 lessons.

Meyer et al. (2002) tested whether 5th-grade students can learn the structure strategy via the Internet with feedback and support from their own personal human tutor. Two approaches to teaching the structure strategy via the Internet were examined. One approach involved an Internet instructor (depicted with a static picture) and retired adults trained in the structure strategy as tutors (displayed with a picture of a tutor or animal representation). Emails from Internet tutors provided delayed feedback on students' last lesson, encouragement, daily assignments, and additional instruction about the strategy with other examples, if necessary. The other approach only involved the Internet instructor, who provided delayed feedback on student work, but no tutors.

The students in the two structure strategy groups were compared to students in a control group who participated in extra sessions of the school's regular reading program. Students were randomly assigned to three groups: structure strategy with tutors, structure strategy without tutors, and a control group. Immediately after the intervention, the groups receiving structure strategy instruction tended to recall more information than the group with extra days of regular classroom reading. The average reader receiving structure strategy training had a total recall score equal to a reader in the control group who scored at the 77th percentile on the immediate posttest (effect size = 0.74). The superiority in total recall for the structure strategy group with tutoring over the control group in reading was clearly evident 2 1/2 months after the end of training. The average reader receiving structure strategy training with the aid of tutors 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). The structure strategy group with help from tutors tended to make more progress in mastering the strategy than the group without tutors. This was particularly the case for students whose messages from tutors focused on providing feedback about the structure strategy and the subject matter of the lessons rather than off-task socializing. Fifth graders in the structure strategy group with tutors made significantly greater gains in self-efficacy than students in the other two groups.

Most students in both structure strategy groups made progress in learning the structure strategy, although few consistently demonstrated mastery of the strategy after training. Those students who mastered the strategy were particularly diligent in their work in the lessons. One English as a second language learner (ELL) was conscientious in completing her lessons and following her particularly skilled tutor's instructions; she made outstanding gains in reading performance. Prior to the intervention she scored at the 28th percentile on a standardized reading comprehension test and did not use the structure strategy. After the intervention, however, she showed mastery of the strategy and scored at the 68th percentile on the standardized reading test. Equivalent texts about different content (supertankers or killer bees) with the same text structure (depicted in Figure 1) were counterbalanced over the pretest and post-test. Her pretest and immediate posttest recall of problem-and-solution texts are listed below (Meyer, 2003):

> *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. On 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" (Meyer, 2003).

This student recalled nine ideas on the pretest about supertanker text shown in Table 2. In contrast to the nine ideas recalled on the pretest, she recalled 88 ideas on the post-test about the problem of killer bees (Meyer, 2003). Most of the students (70%) in the control group did not organize their ideas with a problem part and a solution part. Most completely missed any of the suggested solutions, but this student has both a problem part, signaled with "problem," and a solution part, signaled with "one way." Her last three sentences are an attempt to recall the solution part of the passage. There was considerably more about the solution than produced by fifth-grade students in the control group. In this study we found that a problem with cause(s) and a solution designed to eliminate/reduce the cause(s) was particularly difficult for 5th-grade students.

One important finding from this investigation that was modified in subsequent Web-based structure strategy instructional programs was that 5th-grade students' had difficulty working with a table of signaling words in a pop-up help screen while the text was presented on another screen. A solution to this problem was an instructional aid in the form of laminated keys for each text structure. Each key contained a list of signaling words, an example on the topic of whales, the template/pattern for writing a main idea with the text structure, and a template for writing a recall with the structure. When a text structure was initially introduced to the student, it was added to a key ring available for consultation during the lessons (see Meyer, Wijekumar, & Lin [2011] for a picture of the problem-and-solution key).

One of the challenges that remained in extending this program of research from a paper-based to an electronic medium was the quality of the delivery of the system. Therefore, in more recent investigations, modifications in programming (e.g., use of a pedagogical animated agent to teach the structure strategy) have been explored. In order to increase the accessibility and quality of the delivery of the structure strategy instruction, we developed a Web-based system called Intelligent Tutoring of the Structure Strategy (ITSS) to teach the structure strategy to 5th- and 7th-grade students (Meyer & Wijekumar, 2007).

In several studies the researchers have examined the effectiveness of ITSS with 5th- and 7th-grade students. Unlike the previous Web-based program, ITSS used a web-based delivery system in which several instructional design features could be adapted including: feedback (immediate vs. delayed, minimal vs. elaborated), topic choice (whether or not students could choose text topics for practice lessons), and individualization (the extent of individualization of texts to match students' prior lesson performance).

Meyer et al. (2010) tested whether 5th- and 7th-grade students could learn the structure strategy via ITSS in six months of training for 90 minutes a week spread over two or three days. We examined different feedback and motivation conditions in delivering ITSS with a 2x2 pretest post-test design comparing: a) type of tutor immediate feedback (minimal feedback of "good," & "try again," versus substantial and specific feedback) from the I.T. (Intelligent Tutor, the animated agent); and b) motivational condition (programmed sequence of practice examples vs. student choice of practice examples). In the elaborated feedback version, the animated agent provided elaborated feedback with scaffolding to improve performance on subsequent trials, while the other feedback version involved the animated agent providing only simple feedback about the correctness of a student's response. For example, in response to the same student's performance, the tutor in the elaborated feedback version said, "Your structure, main idea, and details are correct. Great Job! But your signaling words were incorrect. Using your signaling chart (key) as your guide, rewrite the signaling words." On a third deficient trial, students in the elaborated feedback group were given a model response to correct their writing of main ideas. This type of elaborated feedback was not provided for students in the simple feedback condition.

Students were stratified on reading comprehension and then randomly assigned to the conditions (including conditions for counterbalanced testing materials over time of testing). Choice of practice texts affected performance as students completed instruction on the first structure learned – comparison, but had no effect on post-test performances. Number of ideas recalled on this formative evaluation was greater for the choice group, but competency in using the structure strategy was not. Type of tutor feedback affected the ability to identify issues compared when students wrote main ideas in a formative evaluation of the comparison structure, but not the experimenter-designed post-test materials. Below is the pretest recall for a below-average reader in the 5th grade.

> *Pretest* (for an article comparing monkeys): "The monkeys are the smalls Monkeys weghy Less 4 onces a few in. tall."
>
> *Formative post-test* (after 10 ITSS lessons about the comparison structure): "There are 2 different kinds of bats. A Black flying fox bat and a leaf-nosed bat. The Black flying fox bat is one of the bigest, they grow up to 6 feet wide and weigh more than 3 pounds. They are jet black.
>
> Leaf-noised is smaller than the Black flying bat. The leaf-noise bat is only 1 foot wide. The leaf-noise bats come in different (colors) and mostely feeds on masquitoes and moths."

Data from a standardized reading comprehension test (*Gray Silent Reading Test* [GSRT]; Wiederholt & Blalock, 2000) clearly showed that below-grade-level readers made substantial gains from pretest to post-test with all versions of ITSS, and effect sizes ranged from 0.42 to 1.16. All ability levels of readers made greater gains on the GSRT if they received elaborated feedback. Students who received ITSS with elaborated feedback showed more improvement ($d = 0.55$) than students who received ITSS with simple feedback ($d = 0.15$). According to the GSRT manual, the average performance of the simple feedback group corresponds to a percentile rank of 79, while the average performance of the elaborated feedback group corresponds to a percentile rank of 91 (Wiederholt & Blalock, 2000). Substantial effect sizes also were found from pretest to post-test on various measures of reading comprehension, such as recall and strategy competence (d = .39 to .79, Meyer et al., 2010). Students also demonstrated maintenance of performance over summer break on most measures. For example, there was complete maintenance of the ability to use comparative signaling words 4 months after ITSS instruction.

In another design feature study with ITSS (Meyer, Wijekumar, & Lin, 2011), a more individually tailored version was developed to provide remediation or enrichment lessons to better match the needs of 5th-grade readers. Stratified random assignment was employed to compare the effects of two ITSS versions. Fifth-grade students in the more individualized condition made greater improvements from pretest to posttest on the standardized reading comprehension test (d = 0.55) than students in the standard condition (d = 0.30).

Additionally, students receiving more individualized instruction demonstrated higher mastery achievement goals when working in ITSS lessons than students receiving the standard instruction (*d* = 0.53). Also, 5th-grade students receiving more individualized instruction showed greater improvement using signaling, better work in lessons, and more positive posttest attitudes toward computers than students receiving standard instruction. Students in both conditions improved their recall of ideas from texts, use of the structure strategy, and understanding of comparison signaling words.

More recently, 4th- and 5th-grade classrooms have participated in efficacy trials with ITSS conducted in rural and suburban school districts. For this work Wijekumar, Meyer, and Lei (2011) are using a Multi-Site Cluster Randomized Trial (CRT) design to increase statistical power in testing treatment effects. Each school district serves as a site in the Multi-Site CRT design. Within each site or school district, classrooms were randomly assigned to ITSS or a control condition involving the usual language arts curriculum. The study is a replication of past work with 5th-grade students and an extension to 4th-grade students. A modification was made in the instruction to reduce the burden of typing for the younger students. Once students identified two or more topics compared and the issues on which they were compared, they completed a matrix task in which they clicked on text ideas relevant to cells in the matrix, rather than typing the information This procedure to reduce typing was used for both 4th- and 5th-grade students in the first five out of nine practice lessons for the comparison structure. In addition, during implementation some 4th-grade teachers voiced concern about typing demands. As a result, 4th graders, but not 5th graders, were switched to only constructing main idea statements and not recalling all they could remember from the texts in the ITSS lessons. Preliminary results show a significant interaction between grade level and treatment on the post-test with 5th-grade students making more progress. However, 4th graders show significant effects on most measures of reading comprehension but not for details on the comparison text. For the more difficult problem-and-solution text, 4th graders performed significantly better on all measures. This probably results from the fact that without training about the more complex problem with cause and solution, most upper elementary students completely miss the solution part. However, the comparison text contrasted two penguins on a number of the same attributes, many of which were memorable, such as orange ear patches. Fourth graders in control classrooms apparently remembered such isolated details as well 4th graders in the structure strategy classrooms.

In a pretest/post-test design, struggling readers in 4th and 5th grades received special group instruction with ITSS (Meyer & Wijekumar, 2011). The reading teacher who administered ITSS explained that his students, who had the greatest problems in reading, showed a 20 to 70 point gain on the state language arts assessment between 2010 and 2011, which he attributed to the immediate feedback in their structure strategy instruction from I.T. in ITSS. Interestingly, the teacher believed that his 5th-grade students gained more than his 4th-grade students, who seemed to struggle more with ITSS. The data showed statistically significant increases in reading comprehension on a standardized reading comprehension text from pretest to post-test, and no statistically significant time (pretest vs. post-test) by grade interaction (4th vs. 5th). One grade level did not make significantly greater progress than the other. In fact, effect sizes between pretest and post-test reading test scores were .78 and .79, respectively for 4th- and 5th-grade students. Additionally, performance on a signaling test, which tested the ability to use comparative signaling words, also showed statistically significant effects between pretest and post-test and no time by grade interaction. The effect sizes on the signaling test were 1.42 for 4th graders and 1.05 for 5th graders. The fact that 4th-grade students started at a lower level than the 5th-grade students may have led to more difficulty, frustration, or complaining, but they still benefited as much as the 5th-grade students. Nearly all (80%) of the 4th graders in this study used the regular 5th-grade ITSS lessons (constructing recalls) rather the modified 4th-grade lessons (without constructing recalls) used in the ITSS efficacy trial.

ITSS lessons were originally designed for 5th-grade students via matching of standards and curriculum at the 5th-grade level. Perhaps ITSS is best suited to the 5th-grade level in terms of content and readiness for learning about text structure. However, ITSS covers a wide variety of content in science, social studies, history, and some sports topics in order to promote transfer to reading many types of nonfiction texts so interest in ITSS topics should not be confined to 5th grade. Additionally, ITSS may be appropriate for 4th graders because struggling 4th-grade readers had performance gains as large as struggling 5th-grade readers' after using ITSS in a pretest/post-test design (Meyer & Wijekumar, 2011). The greater effects for 5th-grade students in the efficacy trial may have resulted from the better match to the needs of 5th graders or to an ill-advised decision to delete the requirement of creating a recall protocol from the ITSS lessons for 4th grade. The efficacy data cannot differentiate between these two explanations due to different versions of ITSS for the two grade levels.

Currently, another efficacy trial is underway with 7th- and 8th-grade students as part of a grant awarded by the U.S. Department of Education Institute of Education Sciences. The study with 7th-grade students is a replication study of earlier work (Meyer et al., 2010), but also examines effects in both rural and suburban schools. Work with 8th-grade students extends our lessons to older middle school students. In preparation for this extension we wrote lessons to meet Pennsylvania standards in social studies, science, and writing for 8th-grade students. Series of two or three related lessons on a topic using multiple texts with complex structures were added to the compilation of over 100 ITSS lessons as we extended Web-based structure strategy to 8th grade.

Classroom-based interventions in the primary grades. In addition to Web-based instructional programs there has also been renewed interest in teaching text structure in the early grades during the elementary school and preschool years (e.g., Culatta, Hall-Kenyon, Black, 2010; Hall, Sabey, & McClellan, 2005; Hall-Kenyon, & Black, 2010; Williams et al., 2005; Williams et al. 2007; Williams, Stafford, Lauer, Hall, & Pollini, 2009). For example, Culatta, Hall-Keyon, and Black (2010) reported a pretest post-test pilot study project with 71 children ages 4 to 5 years old in four preschool classrooms. The study was a collaborative project of speech-language pathologist and early childhood educators using playful, but systematic instruction in theme-based units over 16 weeks with instruction four days each week. Teachers adapted the unit that was initially co-planned by university researchers (Culatta and Hall-Kenyon). Exit data from participating teachers indicated that the teachers had learned to value expository texts and explicit instruction about them. The intervention focused on two expository text structures: comparison and problem-and-solution.

Meyer (e.g., Meyer & Freedle, 1984) had originally thought that the problem-and-solution structure would be particularly easy for readers due to its similarity to narratives, but learned that was not the case (e.g., Meyer et al., 1989; 2010; Meyer, 2003) when problem-and-solution texts involved identifying and reducing the cause of problems. For example, when adults were asked to identify the most difficult of five text structures to learn and use with the structure strategy, problem-and-solution was most frequently listed. With readers across the life span, use of the structure strategy with scientific text organized with a problem (& cause) and solution (blocking/reducing of the cause) is usually more difficult than text organized with a comparison structure (Meyer et al., 1989). Also, Meyer (2003) reported that over 70% of 5th graders showed no understanding about using the problem-and-solution structure after reading a newspaper article of this type. Overall, research with 4th graders to retired adults suggests that the problem-and-solution texts appear particularly challenging when used in scientific exposition with a problem and its cause(s) and a solution that was aimed at eliminating or reducing the cause(s). Without the underlying causal relationships, this responding structure (question–answer; problem–solution) can be easy to learn, and this appears to be the tact taken by Culatta et al. (2010). A teacher presented a problem in a narrative about a son's escaping hamster to preschool children. Then, the children constructed hamster cages. Next, the problem of hamsters escaping from cages and cages to solve the problem were discussed in an expository framework.

Except for Bartlett's (Bartlett, 1985; 1989; 2010; Bartlett & Meyer, 1981) continued interest in the structure strategy with preschool and elementary school children, until recently there was a dearth of studies with children younger than 5th grade. Williams, Hall, and Lauer (2004) reviewed the literature on awareness of text structure and interventions with text structure, particularly in elementary school grade levels. They noted the lack of intervention studies to teach the structure strategy with students in early grades until recently. They mentioned that a few studies were conducted in the 1980s, but interest in the topic waned without leading to any movements of magnitude in applying research findings to the classroom. However, Williams and her students' research represents major strides in bringing the structure strategy to 2nd grade (e.g., Williams et al., 2005; Williams et al., 2007; Williams, et al., 2009) as well as preschool children (e.g., Culatta et al., 2010).

Williams et al. (2005, 2007) examined direct instruction with the comparison structure (compare-contrast) with second grade students. Her work combines the influences of Meyer (e.g., Meyer et al., 1980; Meyer & Freedle, 1984; Meyer et al., 2002), Armbruster et al. (1987), and her own prior work with narratives, strategies, concept learning, and reading with learning disabled students (e.g., Gersten, Fuchs, Williams, & Baker, 2001). In Williams et al. (2005, 2007) classrooms (128 students in all) were randomly assigned to structure strategy instruction with the comparison text structure, traditional instruction which contained the same content but no structure strategy instruction, or a no instruction control. Signaling words were taught (called clue words); eight clue words from Williams et al. (2005) were "both," "however," "and," "alike," "compare," "but," "than," and "contrast." Note that "and" appears to be a confusing comparison signaling word in that it usually indicates the collection of two things together. Content aligned with the New York State curriculum standards in both language arts and the content areas. Williams' teacher-led instruction with at-risk primary school children used clue words (signaling words), general text structure focused questions (What is this paragraph about? How are they the same? How are they different?), and graphic organizers (e.g., matrix). In addition, the intervention included text analysis (the close analysis of short pieces of well-structured text) and paragraphs that embody the characteristics of a particular text structure. The goal of using well-structured text was to increase familiarity with structure in order to help children strengthen their mental representation of a specific structure. These target paragraphs became more complex across the lessons in the intervention. At first, all the sentences reflect the structure, and later, other sentences (called distractors) are added that include details about the topic but do not reflect the structure. In the second lesson, lions and eagles were contrasted on one issue, skin covering, in a one-paragraph text that also included a similarity (Williams et al., 2005). This contrasts to ITSS where the introductory comparison lessons involved two-paragraph texts that compare two ideas on three contrasted issues. Sequentially moving from one contrasted issue per lesson in early lessons to more contrasted issues in later lessons may be particularly important for at-risk learners and younger readers.

In Williams et al.'s first intervention study (2005) students learned how to classify animals on the basis of issues, such as how they bear their young or get oxygen. Students in the structure strategy group as well as students in the control group learned this content about classifying animals. Adding structure strategy instruction did not reduce learning of vocabulary or other content area concepts. Both groups that received content instruction performed better on the content than the no content control group. Children in the structure strategy group also demonstrated use of the comparison structure with both similar content (animals) as well as new content, indicating that readers were able to transfer strategy knowledge to new contexts (Williams et al., 2005).

Williams, Stafford, Lauer, Hall, and Pollini (2009) replicated these findings in an extended and improved version focused on science topics (e.g., animals) and the comparison text structure. The clue words were reduced from eight to six, and "and" and "than" were deleted from the original list. Additionally, later lessons added a pro-con version of the comparison structure and added some pro-con clue words. As students progress through instructional modules, they become more challenging. Students answer comprehension questions with oral and written responses in early lessons. Later they write summaries with assistance of a form, and by the end of the module they write free summaries without assistance. Again, primary-grade children demonstrated greater learning of the comparison (compare/contrast) structure and improved their ability to understand novel text (Williams et al. 2009). Additionally, they learned domain knowledge similar to students who received content instruction without time devoted to text structure (Williams et al. 2009).

In addition to work with the comparison text structure, researchers have investigated structure strategy instruction with causative text in the primary grades. In a series of lessons developed and evaluated by Williams et al. (2007), the cause and effect structure was taught in the domain of social studies. The lessons focused on the colonists and early American pioneers. The cause-effect general questions were "What happened?" and "Why did that happen?" The clue words were "because," "so," "therefore", and "since." Similar effects were reported as those reported about the comparison lessons and science topics in terms of boosts to reading comprehension and domain learning similar to students who did not learn about text structure. Transfer was observed, but it appeared to be not as strong as that for the comparison lessons (Williams et al. 2007).

In summary, recent interventions with elementary and pre-school students have indicated that the structure strategy can be successfully taught to younger learners, and is associated with improvements in comprehension similar to previous interventions targeted to adults. Although questions remain as to the best method of Web-based instruction with readers in earlier grades, from these interventions several important implications for instruction can be gleaned. First, it is important to provide appropriate scaffolding and instructive feedback to students, increasing the complexity of text and instruction as students improve in their use of the structure strategy. In addition, it is important to select texts which match the reading level of the learner, particularly in initial instruction of a particular text structure. Intervention research also suggests that not only can the structure strategy be taught within a classroom setting; doing so will provide comprehension benefits while not detracting away from content learning. Finally, in adapting structure strategy instruction to younger learners, careful thought of readers' needs and capabilities is needed in order to avoid creating instructional tasks that are too difficult or confusing for young readers to complete.

Conclusions

Much progress has been made in the rigor and extent of research examining the effects of the structure strategy with different types of readers in different contexts. There is substantial and consistent evidence over 30 years that instruction with the structure strategy increases recall from expository text and the organization and quality of readers' recalls. Additionally, there is evidence that structure strategy instruction can increase understanding and use of signaling words, production of good main ideas and summaries, standardized reading comprehension tests scores, and answers to questions. Additionally, structure strategy instruction changes the type of ideas readers underline as important, readers' think-aloud protocols, and their eye movement patterns while reading.

The momentum to bring the structure strategy to elementary school children is currently in full force in part due to funding from the U. S. Department of Education Institute of Education Sciences for the work of Meyer and Williams and their colleagues in recent years. It is hoped that this momentum continues with positive effects on readers from preschool to retirement age. In order for this to happen in K-12 schools, teachers and administrators need to understand the strategy and its importance for reading and learning from text. Although the structure strategy can be provided through a Web-based platform, many students will not acquire the benefits of structure strategy instruction without a school environment valuing and supporting the intervention.

There are still many questions yet to answer about the structure strategy. These include how much instruction to provide with each text structure at different age levels and with what types of texts. Additional research in adapting instruction to meet the needs of children throughout the elementary school years is needed in order to help readers meet the demands of progressively difficult texts in a variety of different domains. Care should be taken so that children learning about text structures in preschool or primary grades realize that they have started on a path of increasing their knowledge and use of the structure strategy with increasingly more varied and complex reading materials and tasks. This tact may avoid giving younger readers the notion that they already know about the structure strategy and need not engage in further instruction. As can be noted in the intervention by Samuels et al. (1988), college students benefited from instruction about the text structure of a scientific journal article. This benefit for reading comprehension held when a journal was presented in the ideal organized manner and when it was not. As readers age, their instructional needs change, and it is important to create structure strategy instruction which is sensitive to the needs of readers in particular phases of their education. Potential areas for study include the timing and amount of structure strategy review throughout years of schooling and extensions to meet the developing needs of the maturing reader. This further research will allow educators to meet the needs of the maturing reader who will interact with more varied and complex informational materials in nonfiction texts and on the Web.

• • •

Acknowledgements

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grants R305G030072 and R305A080133 to The Pennsylvania State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Bonnie J. F. Meyer, Professor of Educational Psychology at The Pennsylvania State University, studies reading comprehension, decision-making, text structure, and the structure strategy. She has over 50 publications about the structure strategy and its effects on understanding and applications for reading or writing. Her current research, funded by the Institute for Education Sciences of the U.S. Department of Education, is focused on structure strategy instruction for students in Grades 4-8. She has served on editorial boards for Discourse Processes, the Journal of Educational Psychology, the

Reading Research Quarterly, Cognition and Instruction, the Educational Psychologist, and the Journal of Literacy Research and is a Fellow in the American Psychological Association, American Educational Research Association, and American Psychological Society.

Melissa N. Ray is a doctoral candidate in Educational Psychology at The Pennsylvania State University. Her primary research interests are reading comprehension, literacy development, and literacy instruction.

References

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, Tempe.

Bartlett, B. J. (1985). Organisational structure: The key to improved comprehension and memory. In L. Unsworth, (Ed.): *Reading: An Australian perspective*. Melbourne: Nelson 1985, pp. 83-102.

Bartlett, B. J. (1989). *Children writing, reading and thinking – and top level structuring.* Unpublished manuscript.

Bartlett, B. J. (2010). Learning about written language, literacy and meaning: A metalinguistic gift . In Raich, M., Schober, P., & Zelger, J. (2010). GABEK IV. Sprachliche Strukturen, Theorie und Anwendungen. Innsbruck: Studienverlag.

Bartlett, B. J., & Meyer, B. J. F. (1981). *A special plan for reading.* Mt. Gravatt Press, Brisbane, Australia.

Berkowitz, S. J. (1986). Effects of instruction in text organization on sixth grade-students' memory of expository reading. *Reading Research Quarterly, 21*(2), 161-178.

Brandt, D. M. (1978). Prior knowledge of the author's schema and the comprehension of prose. (Unpublished doctoral dissertation). Arizona State University, Tempe.

Broer, N. A., Aarnoutse, C. A. J., Kieviet, F. K., & Leeuwe, J. F. J. (2002). The effect of instructing the structural aspect of texts. *Educational Studies, 28*, 213- 238.

Carrell, P. L. (1984). The effects of rhetorical organization on ESL readers. *TESOL Quarterly, 17(4)*, 441-469.

Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly, 19(4)*, 727-752.

Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology*, 80, 448-456. doi:10.1037/0022-0663.80.4.448

Crothers, E. J. (1972). Memory structure and the recall of discourse. In R. O. Freedle & J. B. Carroll (Eds.), *Language comprehension and the acquisition of knowledge* (pp. 247–283). New York: Wiley.

Culatta, B., Hall-Kenyon, K., & Black, S. (2010). Teaching expository comprehension skills in early childhood classrooms. *Topics in Language Disorders, 30*, 323-338.

Dansereau, D. F., Collins, K. W., McDonald, B. A. Holley, C. D., Garland, J.C., Diekhoff, G., et al. (1979). Development and evaluation of an effective learning strategy program. *Journal of Educational Psychology, 71*, 64-73.

Duke, N. K. (2000). 3.6 minutes per day: The scarcity of informational texts in the first grade. *Reading Research Quarterly*, *35*, 202–224.

Dunn, B. R. (1978, March). The use of selected text grammars for describing recall from prose. Paper presented at the Annual meeting of the Southeastern Psychological Association.

Frederiksen, C. H. (1972). Effects of task-induced cognitive operations on comprehension and memory processes. In R. O. Freedle & J. B. Carroll (Eds.), *Language comprehension and the acquisition of knowledge* (pp. 211–246). New York: Wiley.

Frederiksen, C. H. (1975). Representing logical and semantic structures of knowledge acquired from discourse. *Cognitive Psychology, 7*, 371-458

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*(2), 279-320.

Geva, Esther. 1983. Facilitating reading through flowcharting. *Reading Research Quarterly 18*:384-405.

Grimes, J. E. (1975). *The thread of discourse*. The Hague, The Netherlands: Mouton.

Guthrie, J. T., & Davis, M. H. (2003). Motivating struggling readers in middle school through an engagement model of classroom practice. *Reading & Writing Quarterly*, *19*, 59-85.

Hall, K. M., Sabey, B. L., & McClellan, M. (2005). Expository text comprehension: Helping primary-grade teachers use expository texts to full advantage. *Reading Psychology, 26*, 211-234.

Hall-Kenyon, K., & Black, S. (2010). Learning from expository texts: Classroom-based strategies for promoting comprehension and content knowledge in the elementary grades. *Topics in Language Disorders, 30*, 339-339-349.

Holley, C. D, Dansereau, D. F., McDonald, B. A., Garland, J. C., & Collins, K. W. (1979). Evaluation of hierarchical mapping techniques as an aid to prose processing. *Contemporary Educational Psychology*, 4, 227-237.

Jessen, J. (1981). *Effects of reading comprehension on training delinquent adolescents how to identify the top-level structure in text.* (Unpublished master's thesis), Arizona State University, Tempe.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan , J.M. (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior, 14*, 196-214.

León, J. A., & Carretero, M. (1995). Intervention in comprehension and memory strategies: Knowledge and use of text structure. *Learning and Instruction, 5*, 203-220.

Mendoza, L., & Schwarz, A. I. (2011). *Improving comprehension strategies for reading health-related Texts in a Hispanic second language speakers of English population*. Unpublished manuscript.

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, New York.

Meyer, B. J. F. (1974). *The organization of prose and its effects on memory* (Dissertation). Cornell University, Ithaca, New York.

Meyer, B. J. F. (1975). *The organization of prose and its effects on memory.* Amsterdam: North-Holland.

Meyer, B. J. F. (1984). Text dimensions and cognitive processing. In H. Mandl, N. Stein & T. Trabasso (Eds.), *Learning and comprehension of text*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Meyer, B. J. F. (1985). Prose analysis: Purposes, procedures, and problems. In B. K. Britton, & J. Black (Eds.), *Analyzing and understanding expository text* (pp. 11-64, 269 304). Hillsdale, NJ: Erlbaum.

Meyer, B. J. F. (2003). Text coherence and readability. *Topics in Language Disorders, 23*, 204-224.

Meyer, B. J. F., Bartlett, B. J., & Woods, V. (1978). *Facilitative effects of passage structure on recall.* Arizona State University, unpublished paper.

Meyer, B.J.F., Brant, D.H., & Bluth, G.J. (1978, April). *Use of author's textual schema: Key for ninth graders' comprehension.* Paper presented at the annual meeting of the American Educational Research Association, Toronto.

Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of the 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. (1979). *Effects of discourse type on recall* (Prose Learning Series Research Report No. 6). Tempe, Ariz.: Arizona State University.

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., Ireland, J. J., & Ray, M. N. (2011). *Connecting reading and writing: Applying the structure strategy to persuasive texts.* Manuscript submitted for publication.

Meyer, B. J. F., & McConkie, G. W. (1973). What is recalled after hearing a passage? *Journal of Educational Psychology*, *65,* 109-117.

Meyer, B. J. F., Middlemiss, W., Theodorou, E., Brezinski, K. L., McDougall, J., & Bartlett, B. J. (2002). Effects of structure strategy instruction delivered to fifth-grade children using the Internet with and without the aid of older adult tutors. *Journal of Educational Psychology, 94,* 486-519.

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. & Poon, L.W. (2004) Effects of structure strategy training and signaling on recall of text. In Robert B. Ruddell and Norman J. Unrau (Eds.) *Theoretical models and processes of reading*, Fifth Edition, pp. 810-851. Newark, Delaware: International Reading Association.

Meyer, B. J. F., Poon, L. W., Theodorou, E., Talbot, A., & & Brezinski, K. L. (2000, April). E*ffects of adapting reading instruction to the individual differences of older adults.* Paper presented at Cognitive Aging Conference, Atlanta, GA.

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., 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. Pollock (Eds.), *Literacy in African Americans communities* (pp. 233-263). Mahwah, NJ: Erlbaum.

Meyer, B. J. F., Talbot, A. P., Stubblefield, R. A., & Poon, L. W. (1998). Interest and strategies of young and old readers differentially interact with characteristics of texts. *Educational Gerontology, 24,* 747-771.

Meyer, B. J. F., & Wijekumar, K. (2007). A Web-based tutoring system for the structure strategy: Theoretical background, design, and findings. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 347-375). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Meyer, B. J. F., Wijekumar, K., Middlemiss, W., Higley, K., Lei, P., Meier, C., & Spielvogel, J. (2010). Web-based tutoring of the structure strategy with or without elaborated feedback or choice for fifth- and seventh-grade readers. *Reading Research Quarterly, 45,* 62-92.

Meyer, B.J.F., Wijekumar, K.K., Lin, Y. (2011). Individualizing a web-based structure strategy intervention for fifth graders' comprehension of nonfiction. *Journal of Educational Psychology, 103,* 140-168.

Meyer, B. J. F., & Wijekumar, K. K. (2011). *Struggling 4th and 5th grade readers using ITSS*. Manuscript in preparation.

Meyer, B. J. F., Young, C. J., & Bartlett, B. J. (1989). *Memory improved: Enhanced reading comprehension and memory across the life span through strategic text structure.* Hillsdale, NJ: Lawrence Erlbaum.

National Educational Goals Panel (1999). *Reading achievement state by state, 1999.* Washington, DC: U.S. Government Printing Office.

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

Raymond, P. M. (1993). The effects of structure strategy training on recall of expository prose for university students reading French as a second language. *The modern language journal, 77,* 445-458.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376-391). New York: Macmillan.

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.

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.

U.S. Department of Education, Institute of Education Sciences, National Institute for Education Statistics. (2010, March). *The Nation's Report Card: Reading 2009*. Retrieved from http://nces.ed.gov/nationsreportcard/pubs/main2009/2010458.asp

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.

Vincent, J.P. (1985). *Effects of discourse types on memory of prose by young, middle-age, and old adults with average vocabularies* (Unpublished doctoral dissertation), Arizona State University, Tempe, Arizona.

Wiederholt, J. L, & Blalock, G. (2000). *Gray Silent Reading Tests* (GSRT). Austin, TX: PRO-ED.

Wijekumar, K., Meyer, B.J.F., Lei, P. (2011). *Multisite randomized controlled trial examining intelligent tutoring of the structure strategy*. Manuscript in preparation.

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, 2004, 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.

Williams, J. P., Nubla-Kung, A. M., Pollini, S., Stafford, K. B., Garcia, A., & Snyder, A. E. (2007). Teaching cause-effect text structure through social studies content to at-risk second graders. *Journal of Learning Disabilities. Special Issue: Teaching History--in all its Splendid Messiness--to Students with LD: Contemporary Research, 40*, 111-120.

Williams, J. P., Stafford, K. B., Lauer, K. D., Hall, K. M., & Pollini, S. (2009). Embedding reading comprehension training in content-area instruction. *Journal of Educational Psychology, 101*, 1-20.

Yeh, L., Schwartz, A. I., & Baule, A. L. (in press). The impact of text-structure strategy instruction on the text recall and eye-movement patterns of second language English readers. *Psycholinguistic Research on Bilingualism and Reading.*

International Electronic Journal of Elementary Education, 2011, 4(1), 153-168.

Eye-tracking as a tool in process-oriented reading test validation

Oddny Judith SOLHEIM *

Brigham Young University, United States

Per Henning UPPSTAD

Brigham Young University, United States

Abstract

The present paper addresses the continuous need for methodological reflection on how to validate inferences made on the basis of test scores. Validation is a process that requires many lines of evidence. In this article we discuss the potential of eye tracking methodology in process-oriented reading test validation. Methodological considerations are highlighted and special significance is placed on the importance of studying the first reading of a text as well as reading while answering questions about it. This point of view expands the traditional scope of eye-tracking methodology in reading research. We conducted a small-scale study in which 18 12-year olds read and answered questions about a multimodal text. In this study comprehension scores were related to allocation of visual attention in two conditions: (i) reading a text passage for the first time; and (ii) rereading of the text passage while answering questions about it.

Keywords: reading comprehension, assessment, validity, eye-tracking

Introduction

In this article we discuss potential uses of eye-tracking methodology in process-oriented reading test validation. Initially we highlight the challenges researchers face in selecting empirical indicators of reading comprehension and in validating assumptions about what is measured. We then discuss the waypoints where eye-tracking methodology may fortify the empirical foundation for the construction of, research on and validation of reading assessments. Further, methodological considerations are discussed, and results from a small-scale study in which students read and answered questions about a multimodal text are used as example material.

In the 1980s, the psychological tradition of reading comprehension assessment was criticized for taking too little heed of what new, cross-disciplinary insights into reading told us about how readers construct meaning in texts (see Valencia & Pearson, 1987 for a summary of this criticism). Since the 1980s, however, reading assessment has undergone

* Oddny Judith Solheim, National centre of reading education and reading research, University of Stavanger, 4036 Stavanger, Norway. fax: +47 51 83 32 50. E-Mail: oddny.j.solheim@uis.no

ISSN:1307-9298

www.iejee.com

substantial changes in order to integrate new ways of defining reading. Nevertheless, the major challenge in reading comprehension assessment remains the choice of empirical indicators of a theoretical concept which is, in principle, impossible to measure. Since we cannot observe 'comprehension', we must make readers do something to indicate what and how well they have understood (Johnston, 1984). Although many methods of assessment have been explored over this period of time, the most common way of assessing reading comprehension is still to make readers answer questions about a text passage they have read. Answers to questions are indirect measures of reading comprehension that rely on assumptions of what is measured by different types of questions and different item formats. Test items can, for instance, be designed in order to test appointed skills, comprehension processes or reading strategies. One example of this is the four comprehension processes used for item development in the Progress in International Reading Literacy Study (PIRLS – a study to assess reading comprehension in 10-year-olds): (1) focusing on and retrieving explicitly stated information, (2) making straightforward inferences, (3) interpreting and integrating ideas and information, and (4) examining and evaluating content, language and textual elements (Mullis, Kennedy, Martin & Sainsbury, 2006). The recurring discussion of whether multiple-choice and constructed response items test different kinds of literacy and comprehension abilities is also a part of this discussion (Campbell, 2005; Solheim, 2011).

Assessment frameworks often include statements about the kind of comprehension processes or reading strategies that are tested, but when it comes to interpreting score meaning we still face challenges (Messick, 1995; Pearson & Ham, 2005; Solheim & Skaftun, 2009). Messick's (1988, 1995) view of construct validity is based on an integration of any evidence that bears on the interpretation or meaning of test scores, It includes content- and criterion-related evidence as subsumed parts of construct validity. Construct validation is thus a process that requires many lines of evidence. Consistent with this, Campbell (2005) has questioned whether a content focus (text type and topic), common in studying construct representation in educational tests, is sufficient for reading. Campbell claims that reading can be viewed as a unique construct among the many that are typically assessed using educational tests: "it is, perhaps, a construct that has more to do with the process than the accumulation of a knowledge base. As such, questions about the construct representation of reading assessment instruments […] must always include some focus on the cognitive process underlying test-taker performance." (Campbell, 2005, p. 352) This point of view is in line with Johnstons (1984) and Messick's (1995) emphasis on the importance of analysing the processes underpinning item or task performance in educational assessment. Messick defines validity as "an overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy of inferences and actions on the basis of test scores or other modes of assessment" (p.741). Although important, product-oriented test validation offer little detail with respect to the problem-solving approaches used by test takers (Tai, Loehr & Brigham, 2006). In 1930 McAllister (in Afflerbach, 2000) described the problem of indentifying reading process through inferences based on products such as reading test scores. This problem was later described in detail by Alderson (2000) and Allan (1992), among others.

Given the criticism of product oriented test validation, however, the solution to the problem is subject to dispute within the research community. As regards the difficulty of observing reading comprehension, it has been claimed that externalisation of the process of reading may be the only way to make it capable of inspection (e.g. introspection). Several researchers have used introspection methodologies like think-aloud protocols to investigate skills, processes and/or strategies that might be used by students when answering items in reading tests (Allan, 1992; Campbell, 2005; Cordon & Day, 1996; Farr, Pritchard & Smitten, 1990; Langer, 1987; Li, 1992). In think-aloud protocols readers are asked to say whatever they are looking at, thinking of, doing and feeling as they go about solving the tasks. An underlying assumption in this kind of methodology is that participants are in fact aware of the vital features of the process, either during the activity or in retrospect. This assumption has also been questioned. For instance, Afflerbach and Johnston (1984) note that while some participants may not have sufficient metacognitive awareness of their reading processes to report on them, others may have lost the ability to report on processes that have been "automatized". Another objection has been that reporting process might interfere with the reading process, and as such risks distorting and changing the nature of the process (Alderson, 2000; Ericsson & Simon, 1984).

Eye-tracking represents a methodology that gives us an alternative window into comprehension process and problem solving behaviour and can supplement a validation process. By using eye-tracking methodology to externalise parts of the reading process in the form of records of eye-movement patterns, we can explore how comprehension scores are related to actual behaviour. Eye-movement recordings of reading on a discourse level yield an on-line record of the reading process in the form of information about what readers visually focus on in the text passage and for how long they inspect different passages. In an assessment situation eye-tracking data can provide on-line information about readers` decisions to search the text in order to give answer to a question, and about how accurate and effective that search is (e.g., the percentage of time reading relevant information or the use of that information to answer the question).

It has been shown that there is a close relationship between where the eyes gaze and where attention is directed during the processing of visually presented information (see Rayner, 1992), an assumption often referred to as the eye-mind assumption (Just & Carpenter, 1980). The eye mind assumption has been criticized for relying too much on experiments in laboratories without strong support from studies of natural behaviour. Land & Tatler (2009) for instance claim that visual focus in natural behaviour and action is characterized by optimizing focus in order to delimit uncertainties in the total visual scene, rather than a strong cognitive focus. This criticism taken into account, it seems obvious that some reading challenges are more suitable to eye-tracking methodology than others. Multimodal reading material is becoming increasingly important to master (Kress & van Leeuwen, 2001), and it has been claimed that this type of reading should be reflected in assessment instruments to a greater extent (Johnson & Kress, 2003). According to Kamil (2004) new work on multimedia displays has generated renewed interest in eye movements. An understanding of multimodal texts presupposes integration of visual stimuli that are dispersed on a fixed, two dimensional, surface, and therefore require looking at the specific elements. We found reading of this type of text a natural starting point for exploring the potential of eye-tracking in reading test validation.

Eye-tracking methodology has been widely used for reading (for a review see Rayner, 1998) but largely ignored in educational assessment and reading assessment. Paulson and Henry (2002) and Tai et al. (2006) represent important exceptions. Paulson and Henry used eye-tracking methodology in order to investigate claims set forward by the test developers of the Degrees of Reading Power (DRP). The DRP is a commercially available testing program that originally claimed to measure the reading comprehension process, as opposed to the product of reading. The modified cloze setup of the DRP was intended to reflect and measure the process of reading, and Paulson and Henry wanted to explore: "[…] the ways in which students` reading process during the DRP might, or might not, resemble the reading process at work in their reading of materials so that we might evaluate the claims made by the publishers of the DRP" (p. 235). The results of the study showed that the eye movements of participants taking the DRP did not in any way correspond to their reading of unclozed passages. Paulson and Henry concluded that the DRP appeared to cause readers to radically alter their reading process in order to complete the assessment successfully. In 2006, Tai et al. reported results from a pilot study that highlighted problem-solving behaviour during a multiple-choice science assessment. In this study the authors demonstrate how eye-movements differed in a consistent fashion for individuals with known differences in expertise. The study by Paulson and Henry is an example of how we can use eye tracking data to validate assumptions about what we really measure with different assessment tasks. The Tai et al. study is also relevant as it gives an example of how eye tracking methodology can be used to validate that assessment tasks really discriminate between participants at different level of the assessed skill or subject.

An assessment of reading comprehension where students answer questions about a text passage they have read include one or more text passages and a set of test items. This differs from assessments in many other domains where assessments consist of little or no stimuli (for instance text passages or pictures) beyond the test items. The former mentioned science assessment studied by Tai et al. (2006) is an example of this. In such forms of assessments, studying problem solving behaviour during answering is sufficient. However, a focus solely on the problem solving process during answering might turn out to be insufficient when it comes to the most usual form of reading comprehension assessment: reading a text passage and answering questions about it. In this kind of setting, both the first reading of the text and the rereading of it while working with specific questions will contribute to the readers' comprehension of what is read. As such, both of these conditions may represent important sources of information about the reading comprehension process.

Several eye-tracking studies have focused on the initial reading of a text (see Rayner, 1998). In some of these studies observed reading behaviour was related to subsequent measures of recall or reading comprehension (Hannus & Hyönä, 1999; Hyönä, Lorch & Kaakinen, 2002). In eye-tracking studies of this kind the participants are often given unspecified or global purposes for reading (such as having been told to study the text passage in order to be able to answer questions or summarize it). However, several studies have shown that the perspective which is active during reading guides the readers' attentional resources leading to more carefully processing of perspective-relevant information in the text than perspective-irrelevant information (Goetz, Schallert, Reynolds & Radin, 1983; Kaakinen & Hyönä, 2005; Rothkopf & Billington, 1979). In order to investigate the kind of reading behaviour that is associated with successful reading comprehension, we should therefore expand the scope to include the interaction with text while answering specific questions.

In studies of task-oriented reading the interaction with a text passage on a given task is the main focus. Vidal-Abarca, Mãná & Gil (2010) describe task-oriented reading as: " (...) situations in which a reader reads one or more texts while knowing in advance that he or she has to perform a task for which the texts are a crucial and available source of information" (p. 817). They further mention two essential characteristics of readers involved in task-oriented reading: only information that is pertinent to the task performance is relevant to the reader, and the reader interacts with the text on a task, going back and forth from the text to the task and vice versa.

Text availability has often been discussed in connection with reading tests (see, for instance Alderson, 2000). Andreassen & Bråten (2010) found that tests consisting of longer passages answered without access to relevant text increased the predictive power of working memory in reading comprehension scores. Ozuru, Best, Bell, Witherspoon &

McNamara (2007) have suggested that with-text comprehension questions may be more suitable for assessing readers' ability to engage in strategic processing related to specific reading goals. Task-oriented reading, where students have access to the text while searching for information and constructing meaning, is now a common activity in educational settings as well as in assessment settings, and Reading Literacy as it is measured in Progress in International Reading Literacy Study (PIRLS), Program for International Student Assessment (PISA), Assessment of Adult Literacy (ALL) and International Adult Literacy Survey (IALS) is based on text availability.

Some task-oriented reading activities require students to decide what to read first, the task or the texts. If allowed to choose, students also differ in that some students read the questions before they read the text passage, while other students choose to read the text passage first (Salmerón, Vidal-Abarca, Martínez, Maña, Gil & Nauman, 2011). In our study we designed a task in which readers first read the text and then were allowed to search the text to answer the questions. This kind of design has been used in studies of task-oriented reading activities (Vidal-Abarca et al., 2010) and also in studies of reading assessment (Ozuru et al., 2007; Ozuru, Dempsey & McNamara, 2009). In an assessment situation like this we suggest that arriving at a correct answer to a question about the text could be the result of 1) the first general read-through, 2) task oriented reading while answering the particular question or 3) a combination of reading in both these conditions. We therefore split the analysis into two main units: the first reading of the text and task-oriented reading while answering questions.

The stimulus text consisted of both verbal text and illustration. By tracking children's allocation of visual attention we explored what parts of the text passage they read to answer questions that required integration of text and illustration. The underpinning theoretical assumption was that, in order to give a correct answer to the question, the student had to use and integrate information both from the verbal text and from the illustration – in other words, that a correct answer to the question presupposed a given reading behaviour. With a product-oriented approach, a comprehension score would be evidence of the integration of different modalities. By using eye-tracking methodology we related these scores with actual behaviour, and explored to what extent our assumptions about how such tasks have to be solved were reflected in students' eye-movement patterns. We expected that both conditions of reading would have an impact on comprehension of the text, and that students' performance would show diversity in how they exploited the two conditions.

Methods

Participants

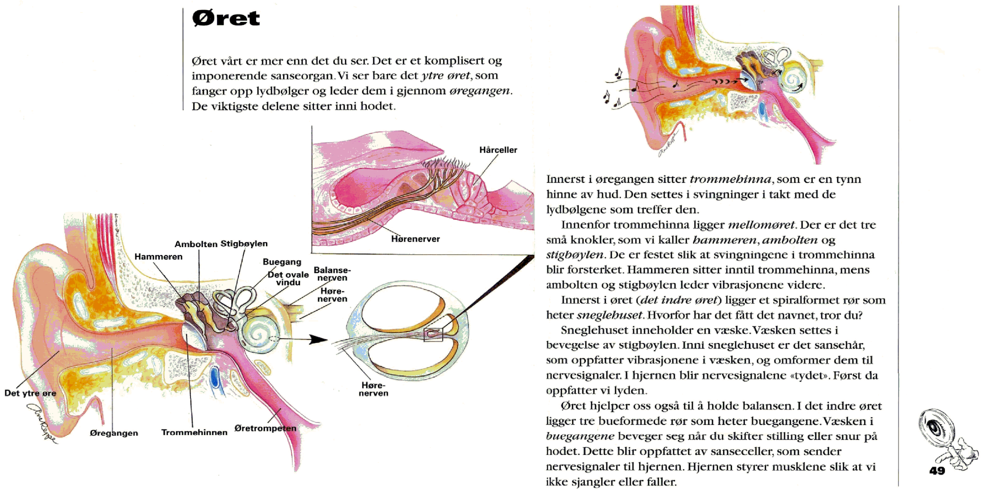
Thirty-four students in seventh grade (mean age 12.75 years) from a medium-sized Norwegian school took part in the study. The sample included 47% female students. We wanted to reduce the likelihood of differences in behaviour and scores being attributable to individual differences in word-reading ability or overall intellectual ability, so 18 students in the centre of the normal distribution on a non-verbal intelligence test, Raven Progressive Matrices (Raven, 1958; Raven, Court & Raven, 1988), and a word-chain test, Ordkjedeprøven (Høien & Tønnessen, 1998) were selected.

Eye-tracking apparatus

The equipment used was iView X HED with a Polhemus headtracker from SMI. This tracker had 50Hz sampling, measuring on the reader's dominant eye. One advantage of this tracker is that readers can move their head and body relatively freely. Gaze position accuracy is $< 0.5°$ (typ.).

Text and questions

The text passage was taken from Globusserien: Naturfag 7 (Johansen & Steineger, 1999), a 7th-year science book. The subject of the passage was 'The Human Ear', and the passage contained both verbal text and illustrations, including captions (see Figure 1). The verbal text described the structure of the ear in a sequential manner, from its outermost to its innermost parts. One of the illustrations was a drawing of the ear where some parts were named in captions. Comprehension of the verbal text was dependent to a large extent on this illustration, and vice versa, but the connection between the text and the illustration was unmarked (the connection was not marked explicitly by means of textual reference, labels or arrows). The reader had to gather both verbally and pictorially presented information from different sections of the page and integrate them into a coherent whole. To accomplish this, the reader had to decide in what order to study the materials – or, in other words, what segments were linked together in terms of content.


Øret

Øret vårt er mer enn det du ser. Det er et komplisert og imponerende sanseorgan. Vi ser bare det *ytre øret*, som fanger opp lydbølger og leder dem i gjennom *øregangen*. De viktigste delene sitter inni hodet.

Innerst i øregangen sitter *trommehinna*, som er en tynn hinne av hud. Den settes i svingninger i takt med de lydbølgene som treffer den.

Innenfor trommehinna ligger *mellomøret*. Der er det tre små knokler, som vi kaller *hammeren*, *ambolten* og *stigbøylen*. De er festet slik at svingningene i trommehinna blir forsterket. Hammeren sitter inntil trommehinna, mens ambolten og stigbøylen leder vibrasjonene videre.

Innerst i øret (*det indre øret*) ligger et spiralformet rør som heter *sneglehuset*. Hvorfor har det fått det navnet, tror du?

Sneglehuset inneholder en væske. Væsken settes i bevegelse av stigbøylen. Inni sneglehuset er det sansehår, som oppfatter vibrasjonene i væsken, og omformer dem til nervesignaler. I hjernen blir nervesignalene «tydet». Først da oppfatter vi lyden.

Øret hjelper oss også til å holde balansen. I det indre øret ligger tre bueformede rør som heter buegangene. Væsken i *buegangene* beveger seg når du skifter stilling eller snur på hodet. Dette blir oppfattet av sanseceller, som sender nervesignaler til hjernen. Hjernen styrer musklene slik at vi ikke sjangler eller faller.

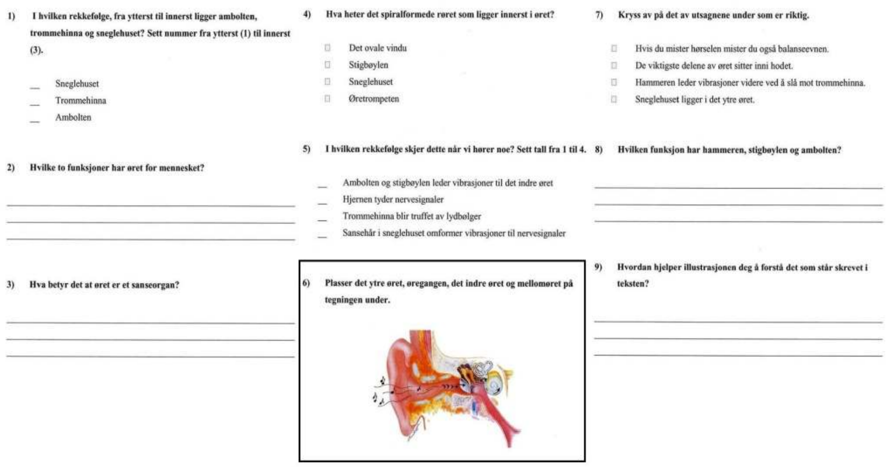
49

Figure 1. The text passage

The assessment was based on a mixture of constructed-response and multiple-choice items (see Figure 2). The analyses of task-oriented reading in condition 2 focused specifically on the pupils' reading in relation to item 6 (see frame in Figure 2). In this item pupils were asked to identify the locations of the outer ear, the ear canal, the middle ear and the inner ear on a drawing.

The item was designed to capture whether the pupils were able to understand and employ a rather complex interaction between text and illustration. The location of the outer ear and the ear canal could be found (marked with captions) in the illustration in the text passage. The middle ear and the outer ear were not named in the illustration, but their locations were mentioned in the verbal text. To identify their locations, the pupils had to combine information from the verbal text with information from the illustration. The text included the following information about the middle ear: "Inside the eardrum is the middle ear. There we find three small bones called the hammer, the anvil and the stirrup."[1] Having read this, the pupil could return to the illustration and locate the middle ear by identifying the location of the eardrum and/or the three named bones. With regard to the inner ear, the following could be read in the text: "In the innermost part of the ear (the inner ear), there is a spiral-shaped tube called the cochlea."[2] To identify the location of the inner ear, the pupils had to identify the cochlea in the illustration. The cochlea was not named in the illustration, but the pupils could identify it from its description as a 'spiral-shaped tube' – or indeed because of the semantic transparency of the Norwegian term sneglehus (literally 'snail-shell').

We assumed that the design of the item and the picture included in the question sheet would direct the participants' attention towards the illustration, but that the information contained in the illustration was not sufficient for answering the question. In fact, the construction of meaning presupposed the combination of visual and verbal information located in different places. To score one point on this item, the pupils had to identify the correct location of at least three of the four parts of the ear. That is, in addition to locating the two parts that could be found in the illustration, they had to locate at least one of the parts that were only mentioned in the verbal text. In parts 2 and 3 of the result section we compare eye-tracking data for participants who answered the item correctly with students who failed to answer this question.


1) I hvilken rekkefølge, fra ytterst til innerst ligger ambolten, trommehinna og sneglehuset? Sett nummer fra ytterst (1) til innerst (3).

__ Sneglehuset
__ Trommehinna
__ Ambolten

2) Hvilke to funksjoner har øret for mennesket?

3) Hva betyr det at øret er et sanseorgan?

4) Hva heter det spiralformede røret som ligger innerst i øret?

□ Det ovale vindu
□ Stigbøylen
□ Sneglehuset
□ Øretrompeten

5) I hvilken rekkefølge skjer dette når vi hører noe? Sett tall fra 1 til 4.

__ Ambolten og stigbøylen leder vibrasjoner til det indre øret
__ Hjernen tyder nervesignaler
__ Trommehinna blir truffet av lydbølger
__ Sansehår i sneglehuset omformer vibrasjoner til nervesignaler

6) Plasser det ytre øret, øregangen, det indre øret og mellomøret på tegningen under.

7) Kryss av på det av utsagnene under som er riktig.

□ Hvis du mister hørselen mister du også balanseevnen.
□ De viktigste delene av øret sitter inni hodet.
□ Hammeren leder vibrasjoner videre ved å slå mot trommehinna.
□ Sneglehuset ligger i det ytre øret.

8) Hvilken funksjon har hammeren, stigbøylen og ambolten?

9) Hvordan hjelper illustrasjonen deg å forstå det som står skrevet i teksten?

Figure 2. The question sheet; item 6 in its own frame

Procedure

Participants were tested during two sessions. In the first one students were tested on the Raven Progressive Matrices (Raven, 1958; Raven, Court & Raven, 1988) and Ordkjedeprøven (Høien & Tønnessen, 1998). Only students who scored between +- 1 SD from the mean according to the standardized norms on both tests participated in the second session, in which the students' eye movements were recorded. Before the actual reading, the eye tracker was calibrated for each participant. The students were verbally instructed to read the textbook passage in order to be able to answer some questions about it afterwards. It was

[1] In Norwegian: 'Innenfor trommehinna ligger mellomøret. Der er det tre små knokler, som vi kaller hammeren, ambolten og stigbøylen.'

[2] In Norwegian: 'Innerst i øret (det indre øret) ligger et spiralformet rør som heter sneglehuset.'

explicitly mentioned that they could look at both the verbal text and the illustrations. The students were given unlimited time to study the passage. When a pupil reported that he or she had finished reading the text passage, he or she was given a question sheet containing a mixture of multiple-choice and constructed-response items. The pupils were given unlimited time to answer the questions. They had access to the text passage and were allowed to look back at it while answering the questions.

Results and Discussion

The presentation of the results is divided into three parts. In Part 1, latent response times and outcome scores for the whole group are presented. The relationship between time spent and total score is also analysed. Parts 2 and 3 use reading behaviour such as it manifests itself through eye movements to investigate relationships between reading behaviour and test performance. In Part 2, gaze duration and integrative saccades are analysed. We investigate the extent to which pupils who answered the item successfully exhibited a reading behaviour different from that of the pupils who did not. In Part 3, we qualitatively explore characteristics of reading behaviour by comparing the first reading with reading while answering the question.

Part 1

The sample was relatively homogeneous with regard to word-reading ability and intellectual ability. However, the data on time spent and total scores show that the pupils differed with regard to how much time they spent and how well they used their time (Table 1). Total score refers to a sumscore based on answers to all the questions on the assessment sheet. Maximum score was 9 points. The participants' total scores varied between 2.5 and 8 points, and the mean score was 5.5. The time spent in Condition 1 (reading the text passage for the first time) varied between 1.3 minutes and 2.9 minutes, with a mean time of 1.9 minutes. The time spent in Condition 2 (answering the questions) varied between 5.4 minutes and 14.9 minutes, with a mean time of 8.3 minutes. There was thus greater variation in how much time the pupils spent on answering the questions (the SD is 29% of the mean) than in how much time they spent on the first reading (the SD is 18.9% of the mean). This indicated that differences in how the pupils constructed meaning would be possible to identify both when they were reading the text for the first time and when they were reading it to answer the questions, which supported the choice of focusing on both conditions.

Table 1. *Outcome score (points) and latent response times (minutes)*

	Mean	SD	Minimum	Maximum
Outcome score	5.5	1.8	2.5	8.0
Response time, condition 1	1.9	0.4	1.3	2.9
Response time, condition 2	8.3	2.2	5.4	14.9
Total response time	10.3	2.2	6.6	16.9

There were no statistically significant correlations (Spearman´s rho) between total score and time spent either in Condition 1 (r = .25, n = 18, p =.31), in Condition 2 (r = .19, n = 18, p = .46) or in Conditions 1 and 2 taken together (r = .20, n = 18, p = .43).In other words, there was no significant relationship between how much time pupils spent on reading the text or on answering the questions and how well they scored on the items. In the analyses in Parts 2 and 3, we will take a closer look at the relationship between reading behaviour as manifested in eye movements and test performance.

Part 2

We divided the pupils into two groups. Item 6 was chosen as the basis for stratification as it demanded a complex reading behaviour (see method section). The '1p Group' consisted of those who scored one point on Item 6 ($n = 8$); and the '0p Group' consisted of those who did not score any points on Item 6 ($n = 10$). The mean total score of the 1p Group was 6.4 ($SD = 1.5$) while the mean total score of the 0p Group was 4.7 ($SD = 1.6$), and this difference was statistically significant ($p < .023$). The aim of this grouping was to explore how scores related to actual behaviour: did pupils in the 1p Group divide their attention between text and illustration differently from the pupils in the 0p Group?

The eye-tracking measurements are based on gaze durations, that is, total time spent on different segments of the textbook passage. Gaze durations were categorised according to areas of interest that were defined around the informational elements (see Figure 3). These areas of interest were the following: the verbal text as a whole (I), verbal text relevant to item 6 (shaded area in I), the illustration as a whole (II) and illustration relevant to Item 6 (shaded area in II). Gaze durations for each of the 18 participants were categorised using the areas of interest in two conditions – Condition 1: First reading of the text passage; and Condition 2: Task oriented reading while answering the question.

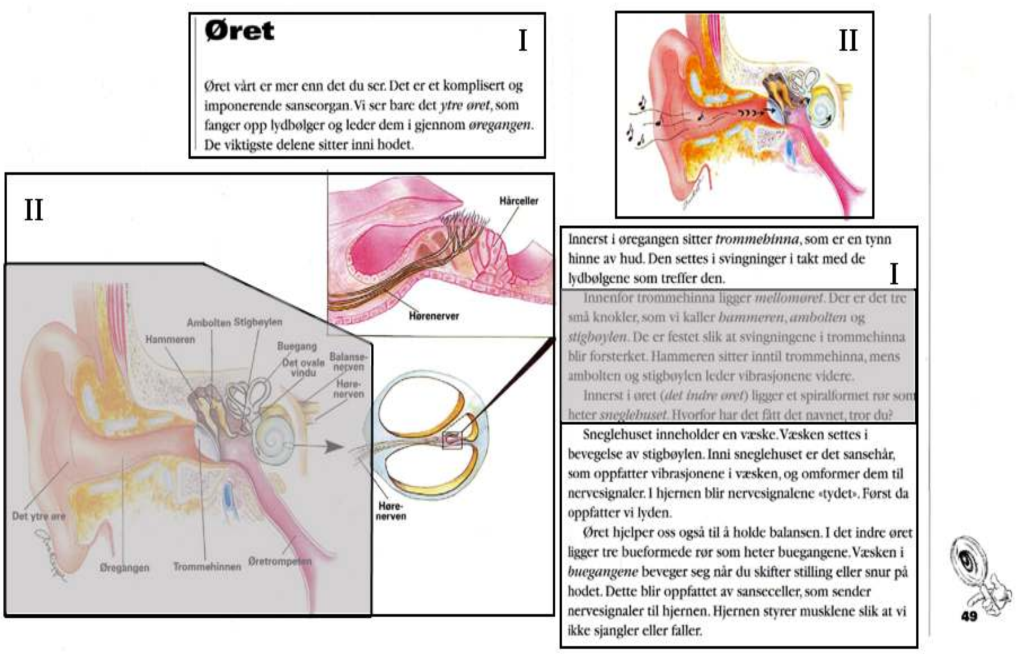


Figure 3. Areas of interest in the analysis. Areas marked I = verbal text, areas marked II = illustration. Shaded areas indicate relevant parts of text and illustration respectively

To score one point on the item, the pupils had to locate correctly at least one of the two parts of the ear that were named only in the verbal text, in addition to the two parts that could be identified using the illustration. We hypothesised that the pupils in the 1p Group divided their attention between the verbal text and the illustration, while those in the 0p Group used mainly the illustration to answer the question. We therefore examined whether there were any differences in how much time the two groups spent on the verbal text while answering the question.

In addition to having paid attention to information both in the verbal text and in the illustration, the 1p Group integrated information from the two modalities. In itself, the fact that a pupil has divided his or her attention between the text and the illustration does not necessarily mean that he or she has integrated the information from the different modalities. Hannus and Hyönä (1999), Holmqvist, Holmberg, Holsanova, Tärning and Engwall (2006) and Holsanova, Holmberg and Holmqvist (2005) have suggested 'integrative saccades' as an indication of bimodal integration. Integrative saccades are transactions between semantically related segments of verbal text and illustration. If there are no integrative saccades, we can assume that the verbal text and the illustration are being read separately. In our study, we defined integrative saccades as saccades following more than one fixation in the text segment of departure and followed by more than one saccade in the text segment of arrival. We examined whether there were any differences between the two groups in the number of integrative saccades made while answering the question. Such reading behaviour may reflect the fact that pupils who answered the question correctly had discovered links between the various parts of the text passage already at their first reading of it. That is why we also investigated if there were any differences between the groups in the number of integrative saccades during the first reading.

Table 2. *Hypotheses and Results: Differences between The 1p Group and the 0p Group. Condition 1= Reading the Text for the First Time, Condition 2= Problem Solving During Answering Questions.*

Condition 1	*Mann-Withney U-test*
Do pupils in the 1p Group have a significantly larger number of integrative saccades between illustration and verbal text than pupils in the 0p Group?	No significant difference between 1p (M = 4.13, SD = 2.75) and 0p (M = 6.4, SD = 5.15); t (18) =, p =.280
Do pupils in the 1p Group have a statistically significantly larger number of integrative saccades between relevant illustration and relevant verbal text than pupils in the 0p Group?	No significant difference between 1p (M = 0, SD = 0) and 0p (M = 1.8 SD = 3.23) t (18) = , p = 3.15

Table 2. *Hypotheses and Results: Differences between The 1p Group and the 0p Group. Condition 1= Reading the Text for the First Time, Condition 2= Problem Solving During Answering Questions.*

Condition 2	*Mann-Withney U-test*
Do pupils in the 1p Group have significantly longer gaze durations on verbal text than pupils in the 0p Group?	No significant difference between 1p (M = 10.62, SD = 17.35) and 0p (M = 6.74, SD = 9.93); t (18), p =.505
Do pupils in the 1p Group have significantly longer gaze durations on relevant verbal text than pupils in the 0p Group?	No significant difference between 1p (M = 7.23, SD = 12.47) and 0p (M = 5.47, SD =9.3); t (18) =, p =.302
Do pupils in the 1p Group have a statistically significantly larger number of integrative saccades between illustration and verbal text than pupils in the 0p Group?	No significant difference between 1p (M = 1.88, SD = 2.59) and 0p (M = 0.80, SD = 1.23); t (18) = , p =.433
Do pupils in the 1p Group have a statistically significantly larger number of integrative saccades between relevant illustration and relevant verbal text than pupils in the 0p Group?	No significant difference between 1p (M = 1.63, SD = 2.26) and 0p (M = 0.60, SD = 1.08); t (18) =, p =.312

As table 2 shows, there were no statistically significant differences for any of these variables between the pupils who scored one point on item 6 and those who did not. Separate analysis of the two conditions showed that some pupils divided their attention equally between the text and the illustration yet obtained different comprehension scores

Part 3

In part 3 we exploit the product measure on item 6 and the process information obtained from part 2 in order to explore the relationship between reading with a global purpose and task oriented reading. Participants were categorised based on their score on item 6 (product) and the extent to which they had used the verbal text while answering the same item (process). The groups were given designations based on qualitative interpretation of characteristic behaviour. The new categorisation yielded four groups:

Task-oriented readers: Pupils who scored 1 point on Item 6 and who read the relevant parts of the verbal text while answering the question (n = 4);

Effortful readers: Pupils who scored 0 points on Item 6 but who read the relevant parts of the verbal text while answering the question (n = 4),

First-time readers: Pupils who scored 1 point on Item 6 but who did not read the relevant parts of the verbal text while answering the question (n = 4),

Non-strategic readers: Pupils who scored 0 points on Item 6 and who did not read the relevant parts of the verbal text while answering the question (n = 6).

Table 3 shows certain similarities between First-time and Non-strategic readers, on one hand, and Task-oriented and Effortful readers on the other hand. The members of the first two groups did not look at the verbal text while answering item 6, but focused instead on the illustration. Common to the pupils in the latter two groups is that they focused on both the verbal text and the illustration while answering item 6.

Table 3. Problem-solving behaviour while answering Item 6. Arrow indicates process, vertical line identifies end of process and asterisk shows obtained product measure.

	1: read question	2: search illustration	3: write part of answer	4: return to illustration	5: read verbal text	6: fulfil correct answer
First-time readers	→	→	→│			*
Non-strategic readers	→	→	→│ →	 →│		
Task-oriented readers	→	→	→	→	→	*
Effortful readers	→	→	→	→	→	

First-time readers and Non-strategic readers

The First-time readers and the Non-strategic readers went through the same steps while answering the question (see Table 3), but yet they ended up with different comprehension scores. None of the pupils in these two groups used information from the verbal text while answering the question, and yet the First-time readers managed to answer the question correctly whereas the Non-strategic readers did not. A closer look at their first reading of the text helped us understand this difference.

First-time readers spent more time on the first reading of the text than the Non-Strategic readers, with an average reading time of 138 seconds versus 108 seconds. First-time readers also distributed their attention to a greater extent across the various parts of the text passage. As a share of total reading time, the pupils in this group devoted, on average, 71.4 per cent to the verbal text and 23.3 per cent to the illustration. By contrast, Non-strategic readers spent, on average, 81.5 per cent of their time on the verbal text and 10.2 per cent on the illustration.

Another interesting measure relates to the share of time devoted to the relevant text. First-time readers spent 50.4 per cent of their total time in the verbal text on relevant text while Non-strategic readers spent 40 per cent. Here, however, it should be noted that the 'area of interest' which we defined in relation to item 6, as 'relevant text' also happens to be the most complex section of the entire text passage. The difference between the two groups may indicate that First-time readers were more likely to adjust their reading speed to the difficulty of the text. At the same time, though, it also indicates that the First-time readers had worked more persistently than Non-strategic ones during their first reading of the text passage, and this thoroughness paid off when they went on to answer the questions.

The eye-movement recordings made during the first reading showed that the two groups had different starting-points when they went on to answer the questions. First-time readers seemed to have laid a foundation during their first reading, which made their problem solving during the answering phase sufficient. We cannot rule out the possibility that First time readers had prior knowledge about the ear that helped them answer the question. Answering the question successfully demands rather detailed and specific knowledge, and if these readers exhibited this kind of prior knowledge it should have been reflected in their total score. However, First time readers did not outperform the other pupils in total score, which led us to assume that it was their reading during condition 1, and not prior knowledge, that led them to the correct answer on this specific item. Non-strategic readers did not have an equally good starting-point when it came to solving the task. It was also characteristic of Non-strategic readers that once they had searched for the answer in the illustration they ran out of solution methods. Some of them gave up quickly when they failed to find the information they were looking for in the illustration, while others continued to search the illustration even though this did not yield any results. The latter pupils were caught in a loop between the question and the illustration, doing something that did not work.

Task-oriented readers

As the First-time readers, the Task-oriented readers managed to answer item 6 successfully, but they solved the task in a different way. While First-time readers scarcely used the verbal text at all, Task-oriented readers devoted on average 16.7 per cent of their "work-with-item 6 time" to the verbal text. The Task-oriented readers used slightly more time to answer the question, and spent a larger share of their time on the text passage (37.6% versus 18.3%). Task-oriented readers used both the verbal text and the illustration, and they had a larger number of integrative saccades and moves between the stimulus sheet and the question sheet than the First-time readers. To sum up, Task-oriented readers seemed to be more active while answering the questions than First-time readers. When we looked at their first reading of the text passage, however, the opposite picture emerged: Task-oriented readers spent less time on their first reading, they spent a smaller share of their time on the illustration (16.3% versus 23.3%), they devoted a smaller share of the total time spent on the verbal text to relevant text (43.5% versus 50.4%), and they had fewer integrative saccades (1.5 versus 5.0). This may indicate that pupils in the two groups 'did the work' at different times. First-time readers were more thorough in their first reading and thus had less need to go back to the text while answering the questions. Task-oriented readers did a somewhat quicker first reading and then let the individual questions govern their further reading.

Effortful readers

The Effortful readers had a response process which was substantially similar to that characteristic of Task-oriented readers. The Effortful readers spent most time of all on answering the question, and they used all the relevant parts of both the verbal text and the illustration while answering. In fact, the Effortful readers 'ought to have had' the information they needed to answer the question correctly, but still they didn't. These pupils exhibited a preparedness in their encounter with the text that indicated that they were capable of finding their way in it, identifying the relevant parts of it and integrating information from its various elements – 'they were almost there'.

The two groups who consisted of pupils who succeeded on item 6 were the groups with the best total scores. The Task-oriented readers had an average total score of 7.5 points, and the First-time readers scored 6.3 points on average. First-time readers made a thorough first reading and seemed to trust in this first reading when they answered the questions. The Task-oriented readers appeared to use their first reading to gain an overview of the text and its various parts, and then let the individual questions govern their further reading. The Non-strategic readers, with 4.3 points, have the lowest total score of all groups. These pupils also used the least (total) time and displayed a limited range of reading strategies. The Effortful readers, whose reading behaviour showed that they worked assiduously and had strategies available to them, but who nevertheless did not manage to exploit these strategies fully, scored 4.8 points on average.

The goal of this example was to explore how eye tracking methodology could add to our understanding of what we measure in reading comprehension tests, and how we should go about to accomplish this. By exploring how students responded to an item that demanded the integration of text and illustration, we could confirm that this particular reading comprehension question, item 6, tap the kind of reading behaviour that we assumed it would. In addition, it also gives us a more nuanced picture of those who did not succeed. Beyond the binary value of right or wrong on the product measure, we get a more nuanced picture of what different pupils need concerning pedagogical approach. This shows a potential usefulness of this type of study in reading test validation.

However, the discovery of the need to analyse both the pupils' first reading and their reading while answering, and comparing behaviour in these two contexts, has important methodological consequences for further research. If we examine only what takes place while participants are working on tasks or while they are reading the text for the first time, we risk drawing the wrong conclusions about the kind of reading behaviour that is associated with solving the task. This should be kept in mind in future eye-tracking studies, but also in studies that use other methodologies, for instance think-alouds. In the current situation no single approach to studying reading process is widely acknowledged. In this situation results should be triangulated with information from complementary methodologies. For instance if we record eye movements during reading and answering, and then replay the recording as basis for a verbal retrospection interview with the reader, we would have an undisturbed recording of reading, and in addition the readers' verbal protocol of what is going on.

Concluding remarks

The pupils who solved the task correctly used both the verbal text and the illustration, but at different times. While some of the pupils 'did the work' during their first reading, others did it while answering the questions. The results presented in Part 3 helped us understand why the approach in Part 2 yielded contradictory information. First, the grouping in Part 2 was based on product measures, and the same comprehension score turned out to conceal very different behavioural patterns. Moreover, each condition was analysed separately. In Part 3, the pupils were divided into groups based on both product and process measures, and their behaviour during the first reading and while answering the questions, respectively, were seen in context. This way of studying the pupils' reading contributed to creating a far more nuanced picture of how different pupils go about constructing meaning in the text than what the approach used in Part 2 was capable of yielding.

In his comprehensive book on reading assessment Alderson (2000) refers to a study by Li (1992). Li used introspection data to show that there were a discrepancy between test constructers' intentions in constructing questions and students reported behaviour in answering them. Li grouped his results as predicted and unpredicted. Predicted results were (i) the expected skill leading to a correct answer; and (ii) unexpected skills leading to a wrong answer. Unpredicted results were (i) the expected skill leading to a wrong answer; and (ii) the unexpected skill leading to a correct answer. If we employ Li´s concepts about predicted and unpredicted results on the results in our study; the predicted results would be: (i) task-oriented readers who looked at both text and illustration during answering leading to a correct answer and (ii) non-strategic readers who only looked at the illustration during reading leading to a wrong answer. The unpredicted results would be: (i) effortful readers who looked at both text and illustration during answering leading to a wrong answer and (ii) first time readers who only looked at the illustration during answering leading to a correct answer. Eye-tracking data from the first reading of the text, however, nuanced these results. A reasonable conclusion then would be that what students do during answering cannot be seen in isolation. The way in which a student has read a text for the first time will probably influence on his or her problem solving behaviour.

Taken together, the combination of product scores and the recording of both reading and answering gives us a stronger empirical foundation for validation of score meaning in reading comprehension assessments. Over the history of science there are numerous examples of how the technology available constitutes new core conceptualizations in a research field. We think that the added information from eye tracking data can get us on the track of understanding more about how students answer questions in reading tests, and enable us to ask new and better questions in future research.

• • •

Oddny Judith Solheim is an Associate Professor in Special Education at the National Centre for Reading Education and Reading Research, University of Stavanger, Norway. Her research focus is on reading comprehension, assessment of reading comprehension, early intervention for children with reading difficulties and reading self-efficacy.

Per Henning Uppstad is an Associate Professor in Literacy Studies at the National Centre for Reading Education and Reading Research at the University of Stavanger, Norway. His research focus is on on-line studies of reading and writing and theoretical work on phonology in dyslexia research.

References

Afflerbach, P. (2000). Verbal Reports and Protocol Analysis. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (Eds.), *Handbook of Reading Research* (vol. 3, pp. 163-179). Mahwah, New Jersey, London: Lawrence Earlbaum Associates.

Afflerbach, P. & Johnston, P. (1984). Research methodology: On the use of verbal reports in reading research. *Journal of Reading Behaviour 16* (307-322).

Alderson, J. C. (2000). *Assessing Reading.* Cambridge: Cambridge University Press.

Allan, A. I. C. G. (1992). *EFL reading comprehension test validation: investigating aspects of process approaches.* Unpublished PhD thesis, Lancaster University, Lancaster.

Andreassen, R. & Bråten, I. (2010). Examining the prediction of reading comprehension on different multiple-choice tests. *Journal of Research in Reading 33* (263-283).

Campbell, J.R. (2005). Single Instruments, Multiple Measures: Considering the Use of Multiple Item Formats to Assess Reading Comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's Reading Comprehension and Assessment* (pp. 347–369). Mahwah, New Jersey: Lawrence Earlbaum Associates.

Cordon, L. A. & Day, J. D. (1996). Strategy use on standardized reading comprehension tests. *Journal of Educational Psychology 88* (288–295).

Ericsson, K. & Simon, K. (1984). *Protocol Analysis: Verbal reports as data.* Cambridge, MA; MIT Press.

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).

Goetz, E.T., Schallert, D. L., Reynolds, R. E. & Radin, D. J. (1983). Reading in perspective: What real cops and pretend burglars look for in a story. *Journal of Educational Psychology 75* (500–510).

Hannus, M. & Hyönä, J. (1999). Utilizations of Illustrations during Learning of Science Textbook Passages Among Low- and High-Ability Children. *Contemporary Educational Psychology 2* (95–123).

Holmqvist, K., Holmberg, N., Holsanova, J., Tärning, J. & Engwall, B. (2006). Reading Information Graphics – Eyetracking Studies with Experimental Conditions. In J. Errea (Ed.), *Malofiej Yearbook of Infographics* (pp. 54–61). Society for News Design (SND-E), Navarra University, Pamplona, Spain.

Holsanova, J., Holmberg, N. & Holmqvist, K. (2005). Tracing Integration of Text and Pictures in Newspaper Reading. *Lund University Cognitive Studies 125.* Lund: Lund University.

Hyönä, J., Lorch, R. F. J., & Kaakinen, J. K. (2002). Individual Differences in Reading to Summarize Expository Text: Evidence from Eye Fixation Patterns. *Journal of Educational Psychology 94* (44–55).

Høien, T. & Tønnesen, G. (1998). *Håndbok til Ordkjedetesten* [Handbook for the word-chain test]. Stavanger: Stiftelsen Dysleksiforskning.

Johansen, E. B. & Steineger, E. (1999). *Globus: Natur og miljøfag 7* [Globus: 7th year science and environment studies']. Oslo: J.W. Cappelen Forlag.

Johnson, D. & Kress, G. (2003). Globalisation, Literacy and Society: redesigning pedagogy and assessment. *Assessment in Education 10* (5-14).

Johnston, P. J. (1984). Assessment in Reading. In P.D. Pearson, R. Barr, M. Kamil and P. Mosenthal (Eds.), *Handbook of Reading Research* (2nd ed., pp. 147–182). New York: Longman.

Just, M. A. & Carpenter, P. A. (1980). A theory of reading. From eye fixations to comprehension. *Psychological Review 87* (329-354).

Kaakinen, J.K. & Hyönä, J. (2005). Perspective Effects on Expository Text Comprehension: Evidence from Think-Aloud Protocols, Eyetracking and Recall. *Discourse Process 40* (239–257).

Kamil, M. L. (2004). The current state of quantitative research. *Reading Research Quarterly 39* (100-107).

Kress, G. & van Leuven, T. (2001). *Multimodal discourse. The modes and media of contemporary communication*. London: Arnold.

Land, M. & Tatler, B. (2009). *Looking and Acting. Vision and Eye Movements in Natural Behaviour.* Oxford, New York: Oxford University Press.

Langer, J. (1987). The construction of meaning and the assessment of comprehension: An analysis of reader performance on standardized test items. In R. O. Freedle & R. P. Duran (Eds.), *Cognitive and linguistic analyses of text performance* (pp. 225-244). Norwood, NJ: Ablex.

Li, W. (1992). *What is a test testing? An investigation of the agreement between students´ test taking processes and test constructers´ presumptions.* Unpublished MA thesis, Lancaster University, Lancaster.

Messick, S. (1988). The Once and Future Issues of Validity: Assessing the Meaning and Consequences of Measurement. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 33–46). Hillsdale, New Jersey, Hove, London: Lawrence Earlbaum Associates.

Messick, S. (1995). Validity of Psychological Assessment: Validation of Inferences From Persons' Responses and Performance as Scientific Inquiry Into Score Meaning. *American Psychologist 50* (741–7499.

Mullis, I. V. S., Kennedy, A. M., Martin, M. O. & Sainsbury, M. (2006). *PIRLS 2006 Assessment Framework and Specifications.* (2nd ed.). Chestnut Hill, MA: Boston College.

Ozuru, Y., Best, R., Bell, C., Witherspoon, A. & McNamara, D. S. (2007). Influence of Question Format and Text Availability on the Assessment of Expository Text Comprehension. *Cognition and Instruction 25* (399-438).

Ozuru, Y., Dempsey, K., McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction 19* (228-242).

Paulson, E.J. & Henry, J. (2002). Does the Degrees of Reading Power Assessment Reflect the Reading Process? An Eye-Movement Examination. *Journal of Adolescent & Adult Literacy 46* (234–244).

Pearson, P. D. & Hamm, D. N. (2005). The Assessment of Reading Comprehension: A Review of Practices – Past, Present and Future. In S. G. Paris & S. A. Stahl (Eds.), *Children's Reading Comprehension and Assessment* (pp. 13–69). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Raven, J.C. (1958). *Standard Progressive Matrices: Sets A, B, C, D & E.* Oxford, UK: Oxford Psychologists Press Ltd.

Raven, J. C., Court, J. H. & Raven, J. (1988). S*tandard Progressive Matrices: 1988 Edition*. London, UK: Oxford Psychologists Press Ltd.

Rayner, K. (1992). *Eye Movements and visual cognition: Scene perception during reading*. New York: Springer-Verlag.

Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin 124* (372–422).

Rothkopf, E. Z. & Billington, M. J. (1979). Goal-guided learning from text: Inferring a descriptive processing model from inspection times and eye movements. *Journal of Educational Psychology 71* (310–327).

Salmerón, L., Vidal-Abarca, E., Mana, A., Martínez, T., Gil, L. & Naumann, J. (submitted). Reading Strategies in task oriented reading: The case of PISA-like tasks. *Manuscript submitted for publication.*

Solheim, O. J. (2011). The Impact of Reading Self-Efficacy and Task Value on Reading Comprehension Scores in Different Item Formats. *Reading Psychology 32* (1-27).

Solheim, O. J. & Skaftun, A. (2009). The problem of semantic openness and constructed response. *Assessment in Education 16* (149-164).

Tai, R. H., Loehr, J. F. & Brigham, F. J. (2006). An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments. *International Journal of Research & Method in Education 29* (185–208).

Valencia, S. W. & Pearson, P. D. (1987). Reading Assessment: Time for a change. *The Reading Teacher 40* (726–732).

Vidal-Abarca, E., Mãná, A. & Gil, L. (2010). Individual Differences for Self-Regulating Task-Oriented Reading Activities. *Journal of Educational Psychology 102* (817-826).

International Electronic Journal of Elementary Education, 2011, 4(1), 169-184.

The emergence of comprehension: A decade of research 2000-2010

Caitlin McMunn DOOLEY*
Georgia State University, United States

Abstract

This review of literature presents research about young children's (ages 2-8) early experiences with comprehension. Using a theoretical framework for emergent comprehension, the review demonstrates how each research study contributes to a holistic theory of emergent comprehension. Influences on emergent comprehension such as children's development, relationships and social interactions, and experiences with multiple texts and multimodal symbol systems are discussed. This review includes contemporary peer-reviewed research articles (spanning the decade from 2000-2010) involving multiple methodologies and representing multiple English-speaking countries.

Keywords: Emergent comprehension, emergent literacy, young children, narrative, literary meaning making, reader response

Introduction

How does comprehension begin? When does a child begin to comprehend text? Ample evidence suggests that children make meaning with texts long before they are conventional "readers." This review of literature presents the theoretical connection between "emergent comprehension" and conventional reading comprehension. Conventional reading comprehension is considered to occur after a child learns to decode (e.g., Gogh & Tunmer, 1986); however, I propose here that comprehension emerges prior to decoding and continues to develop as a child learns to read. To explain the emergence of comprehension, this review highlights children's literary meaning making during early phases of reading development. Throughout this article, I use the terms comprehension, understanding, and meaning making to mean the same thing—to remind us all of the ultimate goal of reading and refrain from piecemeal treatment of comprehension as a set of "strategies" or "skills."

Comprehension, per se, is rarely the topic of study in early childhood literacy even though it is a particularly important topic given children's developmental path toward becoming literate (e.g., Kendou et al., 2005; Paris & Paris, 2003; van den Broek et al., 2005; Wyse & Styles, 2007). A recent meta-analysis by the U.S. National Early Literacy Panel (2008) "present[ed] strong evidence for instructional strategies that promote *code related skills*" (italics added)

* Caitlin McMunn Dooley, Early Childhood Education Department, Georgia State University, Chelsea Circle NE Atlanta, GA 30307, U.S.A. Phone: 404-413-8226. E-mail: cdooley@gsu.edu

ISSN:1307-9298

www.iejee.com

but noted the absence of oral language skills and other meaning-focused research (Lonigan & Shanahan, 2010, p. 344). Similarly, Wyse and Styles (2007) argue against a similar focus on code-related skills in the UK. In the past decade (2000-2010), however, an increasing number of studies have begun to investigate evidence of comprehension among young children, from ages 2 to 8. This review seeks to synthesize this emerging area of research.

Theoretical Framework: Literary Meaning Making and Comprehension

Literary meaning making provides a framework for considering emergent comprehension. Foundational to understanding literary meaning making, Rosenblatt's (1938/1965) transactional theory of reading explains how readers transact with texts to create meaning. Rosenblatt proposed that each reader approaches texts with unique qualities, knowledge, and purposes and strives to construct meanings with texts that may or may not adhere to an author's intended meaning. Rosenblatt suggested that readers seek meanings that are socially compatible—or, in other words, seem to adhere to some constellation of socially constructed meanings; however, no two readers will create exactly the same meaning with the same text. Rosenblatt's theory is now viewed as synonymous with comprehension.

RAND Heuristic for Comprehension

In 2002, the RAND reading study group (RAND Reading Study Group, 2002) defined comprehension by extending Rosenblatt's transactional theory of reading to include the socio-cultural context in which reading occurs. Thus, the RAND group created a heuristic for comprehension, shown below:

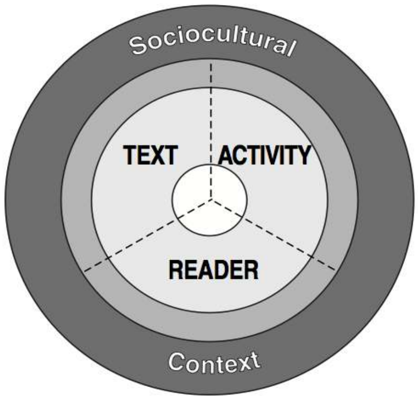


Figure 1. *RAND Heuristic for Reading Comprehension*

Similar to Rosenblatt, the RAND group explained that readers construct meanings in ways unique to their background knowledge, experiences, beliefs, motives, and dispositions. The RAND group also described how texts communicate messages only through readers' meaning making, yet different texts adhere to unique grammars, genres, and lexical qualities. The RAND group also identified the "activity" or purpose for the reading event, a dimension to the heuristic that provides insight into how the immediate context in which reading occurs can guide the purpose for how meaning is constructed. Finally, the RAND group and others have adapted Rosenblatt's theory to include more attention to the socio-cultural context of readers' responses (e.g., Beach, 2000). This added attention to context extends Rosenblatt's focus to include social and political elements that affect the "transaction" yet lie beyond the personal and textual. The RAND heuristic provides a fruitful framework for investigations of comprehension when considering conventional readers.

Emergent Literacy and Emergent Comprehension

As I engaged in research about children's comprehension prior to conventional reading, I found the RAND heuristic lacking. Theories related to emergent literacy helped me to understand that conventional reading frameworks, such as the RAND heuristic, may not fully explain young children's experience. While not all emergent literacy theorists agree on the common elements of children's experiences that inform later literacy, an idea that many emergent literacy theorists share—that literacy emerges over time and with experience—guided my thinking. Clay (1991), the maven of emergent literacy theory, proposed that pre-conventional expressions are not necessarily mimics of adult-like convention. Rather, children express meaning making with texts in ways unique to their development. Children may also engage with text—especially multimodal texts— differently from adult-like convention. In turn, in this literature review, I include studies involving aural, visual, and multimodal comprehension to demonstrate how young children begin to comprehend prior to (or in the beginning phases of) conventional reading.

My colleague Mona Matthews and I adapted the RAND heuristic to demonstrate how emergent literacy theory might inform a study of young children's comprehension. We were moved to do so to explain what we were seeing during a three-year longitudinal naturalistic study of children (ages 2-5) as they learned how to comprehend texts (for further details of this study, see Dooley, 2010; Dooley & Matthews, 2009; Dooley, Matthews, Matthews, & Champion, 2009b; Matthews, Dooley, & Czaplicki, 2011). Dooley and Matthews (2009) presented an adapted version of the RAND heuristic to make explicit how early meaning making extends to later comprehension (see figure below). We call this framework Emergent Comprehension.

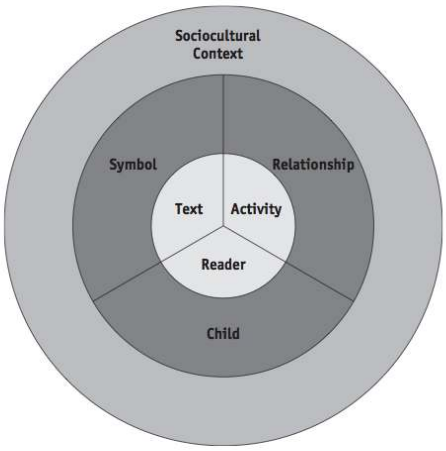


Figure 2. *Emergent Comprehension*

Attending to the uniqueness of early childhood development, this adapted framework identifies the child's cognitive, relational, and symbolic interactions from which meanings are made. The framework contends that (1) children's transactions with text are different from those of older children and adults; (2) young children's symbolic understanding develops across time; (3) meaning making begins at birth through relationships and experiences with important caregivers and peers. Dooley, Matthews, Matthews, and Champion (2009) describe how the youngest learners, from birth through age 2, engage with caregivers to interpret their intentions. As these relationships develop over time, they begin to include text objects (books, computers, phones, etc.) in interactions. Through playful and caring interactions with caregivers and peers, children learn socially acceptable ways to interact with the objects and construct meaning with them (Dooley, 2010). Thus, in the case of printed texts (such as books), a child learns to construct meaning through shared interactions with the object (book) and a caregiver. Over time, the child learns the purpose (or intention) for the object through these shared experiences. If the purpose is to express meaning (such as the case with a book), the child eventually learns to expect that purpose and seek meaning (i.e., comprehend) with the text.

The Emergent Comprehension and RAND heuristic together provide a useful guide for considering early comprehension development as holistic, complex, and emerging with experiences. These frameworks have created an underlying organizational structure for this review. What follows is a brief discussion on the methods for the review and then each study is situated as it relates to the frameworks. In this way, the literature review informs a theory of emergent comprehension so that we can best consider comprehension among young children.

Methods

To conduct this literature review, I began with an online search of multiple ESBCO host databases using terms such as "comprehension," "emergent comprehension," "reader response," "young children," "meaning making," "elementary," and "early childhood." Although oral language is an important factor in comprehension development (Cain & Oakhill, 2007), I only used studies that addressed oral language in the context of "comprehension." After initially finding more than 500 publications, I limited the search results to publications that were written in English, published between 2000-2010, and from "peer reviewed" journal articles. After narrowing to 123 articles, I read titles, abstracts, and skimmed articles to get a summary of their content. I selected articles that involved research with children from ages 2 to 8. Then I categorized articles into groups to identify preliminary themes. I read each article and synthesized how it contributed to, and further developed, the themes. Then, for the purposes of writing this review, I outlined only the articles that seemed to move the field forward in some way. These themes (such as "text talk", "narrative", "multimodality" and others) were then regrouped as they related to the overarching theories of comprehension presented by the RAND and Emergent Comprehension models. After initial reviews of an earlier draft of this manuscript, I also sought particular articles mentioned by the reviewers as relevant to this topic and incorporated these articles in the review. These are the articles I describe here. They present findings from diverse methodological approaches and from international researchers.

Findings

I present the findings in three main sections: (1) Child/Reader: Comprehension as it relates to children's development (2) Symbol/Text: Comprehension as it relates to symbols and texts presented to children, and (3) Relationship/Activity: Comprehension as it relates to the purpose of the social activity at hand. The organization of these findings are intended to provide theoretical groundwork for a holistic understanding of the development of comprehension. However, it is important to note that readers are known to develop at their own pace, in fits and spurts, often with overlapping proficiencies.

Child/Reader: Comprehension as it relates to children's development

When considering comprehension as emergent, we must attend to the uniqueness of children's development. Children's comprehension will look and sound different from adult comprehension. And children's unique cognitive, social, and emotional developmental patterns will inform their ability to comprehend. Research on children's earliest engagements with texts suggest that their "theory of mind," narrative comprehension, and multimodal expressions for engagement are especially important for emergent comprehension.

Theory of Mind. One example of a unique socio-cognitive developmental pattern is a child's "theory of mind." Theory of mind, thought to develop around 2-3 years of age, is an important developmental theory for comprehension because it addresses how humans come to understand that another person can have ideas that differ from their own. For children who hear a story, their theory of mind can affect how they construct the story (and the characters in it). For example, to test whether children's theory of mind affects their interpretation of storybooks, Riggio and Cassidy (2009) presented 45 preschoolers (ages 3-5 years; average age 4:3) with picture books. These books presented characters with "false beliefs"—that is, a belief that was contrary to the reality presented in the book. For example, Little Red Riding Hood had a false belief about the wolf who lay in her grandmother's bed, thinking that it was, indeed, her grandmother. Children who understood the false belief were more likely to focus on the nature of the situation in their retellings (e.g., "she thought her grandmother was there, but really it was a wolf"); whereas, children who did not understand the false belief were more likely to focus on the reality of the situation (e.g., "there was a wolf in her bed") and ignored the false belief altogether. In another study of young children's cognitive development, O'Neill & Shultis (2007) found that children as young as age 3 can understand the differences between a character's physical location and their mental location (where they are thinking about being). These kinds of studies, which account for young children's cognitive development, could provide important details about how children respond to texts in ways unique to their earliest life phases.

Narrative Knowledge. Children's knowledge and understanding of narrative is also a cognitive development that informs comprehension. Considered a "top-down" skill, children's abilities to understand and retell stories, to recognize characters, and to identify causal elements within narrative have garnered recent attention from researchers. In their validation of a narrative comprehension assessment for children ages 4-6, Paris and Paris (2003) confirmed the developmental nature of children's ability to create narratives from pictures. Likewise, Lynch, van den Broek, Kremer, Kendeou, White, and Lorch (2008) demonstrated that between ages 4 to 6 years, children show increasing sensitivity to narrative elements (e.g., coherent episodes, character knowledge, and causal connections), thus supporting the hypothesis that narrative knowledge develops over time among young children. They also suggested that narrative knowledge correlated with the children's story recall, ability to identify overall causal structures (for age 4 only), word identification (age 6 only), and ability to answer comprehension questions about a text that was read to them or shown via video. Alternately, the researchers did not find any correlation between the children's narrative knowledge and phonological awareness or letter identification. Thus, the researchers theorize that narrative comprehension develops separately (although concurrently and complementarily) to other reading skills.

Multimodal Expressions. Attention to the uniqueness of children's development requires that researchers attend to young children's multimodal expressions and representations of story worlds (e.g., Dooley, 2010; Siegel, 2006; Siegel, Kontovourki, Schmeir, & Enriquez, 2008; Wohlwend, 2008a, 2008b, 2009a, 2009b, 2009c, 2011). This shift is both pragmatic, because the youngest learners may not yet be able to verbalize their understandings, and inclusive because attending to multimodal expressions provides an "enabling framework" (Flewitt, Nind, & Payler, 2009). After studying the classroom, home, and playgroup experiences of a four-year-old child with learning difficulties, Flewitt, Nind, and Paylor (2009) suggest that adults seek to understand children's idiosyncratic modes of communication as attempts at intentional meaning making, regardless of ability/disability status. They describe the "politics of semiotic resources/modes, where the established priority given to the conventions of spoken and written language have an impact on what different settings offer [to children with learning disabilities]" (p. 231).

Attention to children's unique developmental patterns should inform research and practices that involve emergent comprehension. Young children's emergent comprehension experiences are likely to emerge from playful and/or caring experiences and will be influenced by their social, cognitive, and emotional qualities.

Relationship/Activity: Comprehension as it relates to the social activity at hand

In the past decade, researchers have tried to pinpoint what aspects of these activities are most influential, and the interactions, listening, and responsiveness of talk seems to be especially important. Thus, we can surmise that these are indicators of a responsive child-caregiver relationship that lays the foundation for comprehension.

Importance of Relationship. Most studies of children's earliest experiences with comprehension involve them listening to a caregiver's stories and talking about those stories. Interestingly, these stories do not have to come from storybooks at all to be beneficial, they simply can be elaborative stories told by the caregiver (Reese, Leyva, Sparks, & Grolnick, 2010). This begs the question: How important is it that the storyteller is a caring adult? Does a caring relationship matter? Studies demonstrating very young children's (ages 0-2) inability to learn from television suggest that caring adult-child relationships and responsiveness do matter for very young children's meaning making (American Academy of Pediatrics, 2011). Dooley, Matthews, Matthews, & Champion (2009) suggest that through caring relationships early in a child's life, the child becomes cognizant of the meaning making affordances of texts. Because young children have a strong need to belong (called the "basic affiliative need"), they are likely to engage in activities and learn how to make meaning in ways promoted by the caring adult. Eventually, these early adult-child relationships extend to include talk, texts, and other symbols and objects. Such is the case in Reese, Leyva, Sparks, and Grolnick's (2010) study: As their mothers reminisced, the children gained important knowledge about narratives.

Adult-Child Talk. A majority of investigations of emergent comprehension have become attentive to specific characteristics of adult-child talk. Studies of literary meaning making with preschool children suggest that children benefit from adults' skillful orchestration of talk, turn-taking, and timing. The cognitive "level" of talk (e.g., Price, van Kleeck, &Huberty, 2009), the number and length of utterances (e.g., Taylor & Pearson, 2004), and narrative qualities such as the coherence of sequenced events, characterization, macrostructural devices such as theme development, and microstructural devices such as story-narration language (e.g., "once upon a time…", "suddenly") (e.g., Epstein & Phillips, 2009) have become part of detailed analyses of children's responses to texts. Adult talk is analyzed mostly for its interactivity, thus researchers are most attentive to the number of conversational turns, the types of questions or prompts adults pose, and the types of children's responses elicited (e.g., Zimmerman, Gilkerson, Richards, Chistakis, Xu, Gray, & Yapanel, 2009). These in-depth descriptions allow us to witness the child/caregiver talk through a complex lens.

We can conclude that children mimic the type of talk that adults model; however, more is not always better when it comes to adult-child talk. Price, van Kleeck, and Huberty (2009) compared the talk between 62 dyads of parents and their 3-4 year-old children. They looked at the amount of talk, syntactic complexity and vocabulary of that talk, the content of the talk across two types of text genres: expository and narrative. They found that expository texts elicited more extratextual talk by children and adults, and when adults talked more, children did as well. However, the content of the vocabulary and concepts did not rise with more talk. Alternatively, Zimmerman, Gilkerson, Richards, Christakis, Xu, Gray, and Yapanel (2009) studied 71 families with children ages 2-4 over an 18-month period. After studying extratextual talk between children and adults, they found that the adult's ability *to listen* to a child's response and respond conversationally does indeed increase the amount and quality of children's talk.

The usefulness of adult/child interactions and talk for insight into children's comprehension continues into the early elementary grades. As researchers grapple with how to distinguish primary-grade children's ability to comprehend prior to (or in the earliest phases of) conventional reading, they also have attempted to categorize text talk and responsive interactions. McKeown and Beck (2006) suggest that teachers need to follow-up read alouds with "prompts for elaboration and completion" (p. 286) and ask "open questions" that stretch children's responses (as opposed to "constrained questions" that allow for single-word responses like yes or no). Sipe has been an influential guide to defining how primary-grade children express their own literary understanding (Sipe, 2000; 2008). Sipe presented a grounded theory that includes five aspects of literary understanding: (1) Analytical (children analyze the single text as a cultural object); (2) intertextual (children relate across texts); (3) personal (children relate texts to their own lives); (4) transparent (children engage in the story world presented in a text); and (5) performative (children manipulate story for their own purposes, usually playful; also described in detail by Adomat, 2010). Sipe (2006) added another aspect: resistance. Children expressed opposition to texts because of conflicts between their known stories and the stories presented, objections to the author's craft, and evocation of painful realities. These categories have provided a useful framework for understanding young children's complex approaches to comprehending stories. And while other researchers have also categorized children's talk around books, we seemed to have hit a point of saturation in this kind of categorical research. Researchers are now seeking to explain how different aspects of meaning making might be elicited.

Interactive Play. In addition to the quality of talk, the importance of interactive play cannot be underestimated as inquiry about young children's meaning making stretches to include young children. Landry and Smith (2006) argue that although research evidence suggests that parents' interactive strategies assist with language development, "less is known about relations between the home literacy environment and aspects of early literacy skills other than language" (p. 136). This is an area ripe for research. For example, Matthews, Dooley, and Czaplicki (2011) have demonstrated that parent/child interactions involving play can inform emergent comprehension. Likewise, Roskos, Christie, Widman, and Holding (2010) summarized three decades of research (192 studies) that connects early play experiences to literacy learning. Sixty-seven percent of the studies showed modest to large effects between play and the "early literacy comprehension domain" (p. 74). They concluded that pretend play was an essential pathway for children to attain "theory of mind," an essential understanding undergirding literacy development. Recently, Wohlwend (2008, 2009a, 2009b, 2011) has researched how children construct, respond to, and represent story worlds through interactive play. Wohlwend's studies extend our understandings by demonstrating how children's early attempts at literary meaning making are mediated by their social relationships. These studies demonstrate how interactive play support literary meaning making and are essential to emergent comprehension. Indeed, there is a need for research on interactions that extend literary meaning making to play and other modalities.

Symbol/Text: Comprehension as it relates to symbol systems and texts presented

Over the past decade, perhaps the most striking new direction in research on comprehension is about the growing diversity of texts. An emergent comprehension framework suggests that printed texts rely on one kind of symbol system: Print. There are other symbol systems such as images and oral language that come from modes besides print. Kress (2010) proposes that a semiotic theory of comprehension is especially timely because of the increased accessibility of multimodal texts and the increase in global communication. A semiotic view of comprehension requires attention to multiple modes and symbol (or sign) systems, printed text being just one symbol system of many. Thus, the term "symbol" is used in Emergent Comprehension to allow for multiple systems, including but not limited to printed text.

Palinscar and Duke (2004) suggest that the "explosion of texts" (p. 183) now accessible via the web, the availability of hypertext and hypermedia, and a renewed attention to non-fiction in the early grades have changed the ways in which researchers investigate young children's meaning making. Likewise, Bearne (2003) suggests that the field of literacy studies is undergoing a "paradigm shift" because children today are exposed to an environment full of multimodal digital texts. These newly accessible texts have created a greater array of genres to choose from, blurred genre boundaries, and prompted new questions about how teachers and families can mediate young children's learning. Researchers increasingly have documented how children construct meaning with multimodal texts, informational texts, e-texts, and postmodern picture books. Each text type elicits unique conclusions about emergent comprehension, described here.

Multimodality and Symbol Systems. Comprehension emerges via exposure to multiple kinds of texts, not just printed texts; thus, we can use the term "symbol systems" to describe the differences in texts. Symbol systems are meaning making systems. They include images or printed text or animation or sound or aural language or gesture (to name a few). Or symbol systems can incorporate and synchronize across these modes. One could argue that the picturebook has been a long-standing multimodal text; however, only recently have researchers been attuned to the uniqueness of multimodal texts as a means for early comprehension development.

The Simple View of Reading (Gough & Tumner, 1986) was perhaps one of the first multimodal theories of comprehension. The Simple View proposed that children's oral language skills (plus their decoding skills) influenced comprehension. Oral language has been shown to independently predict children's (ages 4-6) reading comprehension two years later (Cain & Oakhill, 2007; Kendeou, Savage, & van den Broek, 2009; Kendeou, van den Broek, White, & Lynch, 2009).

Yet current research takes the Simple View further to suggest that not only oral language can contribute to comprehension, but also that exposure to other modes can contribute, such as looking at pictures and television viewing/listening (Evans, 2009; Kendou, Lynch, van den Broek, Espin, White, & Kremer, 2005; Lynch, van den Broek, Kremer, Kendeou, White, & Lorch, 2008). In addition, over the past decade, researchers are questioning the seemingly sequential perspective of the Simple View (i.e., that a child needs language skills before comprehension can ensue). For example, Lynch et al. (2008) propose that comprehension skills, such as inference making, identifying causal relationships, and sequencing, can emerge from children's experiences with aural and visual texts. Indeed, "comprehension skills develop simultaneously with, rather than following, basic language skills" (Kendou et al., 2005, p. 91). Likewise, Paris and Paris (2003) confirmed (in their validation of a narrative comprehension assessment for children ages 4-6) that the children's comprehension of narratives was consistent with their comprehension in other media (e.g., videos). Comprehension seems to emerge with language as well as other modes for meaning making.

Walsh (2010) contends that multiple modes create different cognitive pathways and processes. Walsh's research compares how children in early primary grades "read" print-based texts and multimodal texts. She found that children reading visual texts in kindergarten (ages 5-6) are more likely to observe, label, and comment. Whereas, in Year One (ages 6-7) children reading visual texts are more likely to create intertextual connections and offer personal opinions that are more evaluative. Investigating how children in primary grades (Year One [ages 6-7] and Year Three [ages 8-9]) read digital texts, Walsh (2007) found that children engaged in different patterns of response. They were less likely to engage in critical reading, evaluating, or inferring. Walsh questions whether children's reading practices for one mode transfer completely to another mode. Thus, we are left to wonder whether and how comprehension skills might generalize across modes (as Kendeou, Bohn-Gettler, White, and van den Broek [2008] suggest) and when transfer of comprehension from one mode to another is unlikely (as Walsh [2007] suggests).

Postmodern Picturebooks. Children's literary meaning making with postmodern picture books may be important windows to understanding how they navigate the intricacies of multimodal texts. Postmodern picture books present stories that are often comprised of multistranded narratives (from different characters' perspectives and/or from different points in time). They are often non-linear, with non-sequential plots. The narrators of these stories often address the reader directly and comment on their own narration. Intertextuality—a combination of narratives within a text—is common to postmodern literature. And parody—mimicking conventional texts but including a critical difference—often characterizes postmodern literature (Pantaleo, 2004; Galda & Cullinan, 2002). Postmodern children's literature presents unique opportunities for children's responses because of its complexity.

Two important studies demonstrate children's responses to postmodern literature (Arizpe, 2001; Pantaleo, 2004). The youngest children studied (ages 6-7) responded to Anthony Browne's *Voices in the Park* (2001) (Pantaleo, 2004). Browne's picture storybook presents four intertextual narratives from the perspective of four visitors to a city park. The trajectory of events is non-linear and non-sequential. And the perspectives narrated present a critical view of economic and other social boundaries that segregate the characters. One of nine books used over 10 weeks in Pantaleo's study, *Voices in the Park* was shared with the first grade class over seven read-aloud sessions. Pantaleo found that the children strived to make intratextual connections, or "text-within-the-same-text connections" (p. 220). Yet the children mostly did not notice or comment on the parodies or allusions within the text (e.g., the human mother's wolf-shaped shadow; the beggar dressed as Santa Claus). They engaged in interpreting the synergy between images and text presented in the book, often beginning a read aloud session by searching the images for "secrets" or visual jokes prior to allowing Pantaleo to begin reading. Pantaleo used Sipe's (2000) categories to characterize the children's responses and found that they fit well.

Arizpe (2001) studied the responses of children (ages 8-9) to Anthony Browne's *The Tunnel* (1990/1997). This book presents vivid images and scant text to tell a story of Rose, a bookish girl, who crawled through a tunnel to follow her brother, only to find him turned to stone. When she hugs him and her tears fall upon him, he is brought back to life. Arizpe (2001) shared the book with 72 children from three schools in London and Cambridge, UK. She found that the children searched for narrative structure among the pictures. And their familiarity with the text (after repeated readings) increased the depth and dimension to their responses. Interestingly, however, Azripe analyzed children's responses to the visual images and found that children searched for meaning or reference for any perceived symbol. In other words, they sought to interpret visual referents, even unconventional referents, as they searched for narrative structure. She also used background knowledge of fairytale genre to piece together the events created through image into a fairy tale. This macro-structural genre knowledge enabled their literary meaning making.

Multimodal (and postmodern) texts necessitate an expansion of interpretive repertoires. Pantaleo suggested that the "synergy among the various metafictive devices in *Voices in the Park* creates an overarching indeterminacy in the text and positions the readers in a co-authoring role" (p. 226). Her point is well-taken with respect to postmodern and multimodal texts. As children, and other readers, search for converging meanings among multiple modes and intertextual narratives, they are positioned more and more as active meaning makers. Serafini (2010) suggests that these dynamic texts require new research and pedagogy to explore how to broaden the methods and perspectives to guide young readers to stitch together meaning.

Informational Texts. Informational texts have gotten much attention lately, perhaps in response to Duke and colleagues' eye-opening study of the lack of such texts, especially in early elementary classrooms serving low-income communities (Duke, 2000). Since then, Duke and colleagues designed a quasi-experimental study to determine whether access to informational texts was beneficial to first graders. They found that "there was no harm, and some modest benefit…in including more informational texts in children's classroom environment and reading and writing activities" (Palinscar & Duke, 2004, p. 189). Duke and colleagues also found that access to informational texts did not impede children's attitudes toward narrative texts and they were just as able as the comparison group to write narrative texts. These results suggest that children's interactions involving informational texts *benefit* literary meaning making. Additionally, Smolkin and Donovan (2001) reviewed teacher-student discourse in a first grade class (ages 6-7) over two years. Eighty-three percent of that talk was "meaning oriented" when students were using informational texts. Only 16% of the talk was "meaning oriented" when they were using storybooks (in contrast to, for example, procedural statements like "turn the page" or "I can't see"). Thus, interactions around informational texts may provide opportunities for children to work on comprehending text meanings. Intuitively, we believe that improved comprehension would strengthen children's literary meaning making, because they are different dimensions of the same construct (i.e., meaning making with texts). Future investigations could explore what combination of informational and storybook texts is most useful for children as they learn to make literary meanings as well as efferent meanings.

E-Texts. The newest multimodal text-types available to young children are e-texts (also called digital texts), such as e-books available on computers, digital reading devices, and other platforms. Digital texts can transform reading, and thus, meaning making, because of their capacity to integrate multiple modes by which a reader will construct meaning (Korat & Shamir, 2006). These texts often feature special tools that allow for inserting notes, highlighting and bookmarking text options, retrieving of dictionary meanings, adjusting font size, translating text-to-speech, and following hyperlinks to related texts. Korat and Shamir (2006) compared children's (ages 5-6) emergent literacy levels before and after engaging with e-books. They varied the activity modes accessible to the children: "read story only," "read with dictionary," and "read and play." Examination of pre- and post-intervention measures showed that children in the "read with dictionary" and "read and play" activity modes improved more emergent literacy skills than the "read story only" mode. Thus, it seems that exposure to digital integration of multiple modes on an e-book can result in greater understanding of the story.

Perhaps because of their newness, technologies related to early childhood literacy development are vastly under-researched. A review of the four English-language journals with the highest circulation rates in language and literacy studies (*Reading Research Quarterly*, *Journal of Literacy Research*, *Written Communication*, and *Research in the Teaching of English*) revealed "extreme marginalization" of research on digital tools in early childhood education (Lankshear & Knobel, 2003, p. 64). Lankshear and Knobel point out that of the studies that exist, most present digital tools that emphasize stand-alone, non-interactive media (i.e., not networked) and focus on decoding skills and basic writing/handwriting skills. Rarely do studies of digital tools focus on interpretive forms of meaning making. Thus, this is truly a "great frontier" for research on literary meaning making.

Most relevant to young children's literary meaning making, CD-ROM talking books have been investigated as potentially beneficial texts for reader response. Labbo and Kuhn (2000) investigated a young child's understanding of story after interacting with a CD-ROM storybook and found that the multimodal features could become "inconsiderate" to young readers. These features distracted children when they were incongruent or incidental to the story. Their findings converge with Bus & Neuman's (2009) recommendation that digital texts present stories in coherent ways so that children can retell the story and respond to the text as a story. They warn that the "extra" games and animations that accompany "Living Books" sometimes fragment the story simply through gadgetry and cause children to become passive witnesses to these text-like novelties rather than engage them in the story world.

More recently, Larson (2010) studied how 17 children in a second grade classroom (ages 7-8) responded to reading using a Kindle digital reading device. She focused her analysis on observations of two girls who partnered to read *Friendship According to Humphrey* by Betty G. Birney (2006) for 40 minutes daily. Larson found that both girls enjoyed using the digital devices to read and the digital note-taking devices prompted their responses to reading. Larson categorized their responses as: (1) understanding of story (e.g., retelling); (2) personal meaning making; (3) questioning; (4) answering questions posed in the text; and (5) response to literary features/literary evaluation. Both girls engaged in all types of responses. Most of their responses were in categories 1 (understanding the story) and 2 (personal meaning making). Both girls reported that they enjoyed reading the Kindle e-book more than hard-copy books. Neither liked the "text-to-speech" feature because of the computer-like voice. However, the student described as having a lower reading ability used the dictionary and text-to-speech feature to help her decode words and difficult passages. Larson noted that this student and her parent also reported that she felt more confident as a reader when using the Kindle.

Larson's study prompts us to wonder what e-books offer that hard-back books do not? Does digital integration within a text make a difference for comprehension? Are the devices that carry these texts just as important as the digital nature of the texts? Each platform presents different affordances and constraints, yet none are yet so widely available in schools that they have transformed reading. Indeed, many children are still hoping for access to *20th Century* literacy tools—books—much less 21st Century tools. Yet, we hope that tools such as the iPad, Nook, Kindle, SonyReader, or even the FisherPriceiXL Learning System become more readily available to classrooms simply because of the "coolness" factor. Perhaps digital books could entice kids to read more and better, much like word processors have enticed them to write more and better (e.g., Shamir, 2009).

Discussion

Literary meaning making is indeed one of the most important experiences contributing to emergent comprehension (Dooley, 2010; Dooley, Martinez, & Roser, in press; Dooley & Matthews, 2009a). In many countries, children engage with stories and informational texts early in life, on the laps and by the sides of caring adults. These repeated engagements build familiarity of narrative while also enticing children into the "story world" that will continue to engage their interests and build their knowledge about the world. From the time a child begins hearing and seeing stories, whether told from a family member or viewed on television or stitched from images, that child begins to construct narratives. Children's responses to those stories, whether represented through play, or talk, or image, both demonstrate their understanding while also strengthening the child's meaning making. Over the past decade, research on comprehension among very young children has demonstrated that children not only *can* understand texts but that these understandings endure. Future investigations might inquire how children's unique social, emotional, and cognitive developmental patterns inform their comprehension. And, following the lead of Wohlwend (2011) and others, we might ask: If we know that play and close relationships enhance emergent comprehension, how can we leverage these understandings to improve family support, child care, and early childhood education?

Research on children's multimodal responses as well as their transactions with multimodal texts such as postmodern picture books, television shows, and digital texts suggest that comprehension can be constructed with media beyond print. Thus, we can conclude that emergent comprehension is a learning process that occurs prior to (and during) beginning conventional reading. While there is some indication that children generalize across modes as they learn to comprehend (e.g., Kendeou, Bohn-Getter, White, & van den Broek, 2008), there is also indication that comprehension is likely to be influenced by one mode more than another at different phases of development (Tilstra, McMaster, van den Broek, Kendeou, & Rapp, 2009; Walsh, 2010). The question remains: In what ways do multiple modes contribute to comprehension as a child becomes a more conventional reader? And how might emergent comprehension (and expressions of comprehension) via multiple modes enhance children's conventional reading comprehension? Are some modes more informative than others? Do the contributions of different modes vary with time and development?

The next decade promises improvements in research, pedagogy, and practice. We hope for improved understandings of the effects of early development on young children's progress as literary meaning makers; improved access to various text types for children in schools and homes, especially for low-income communities; improved performance assessments that teachers can use to gauge instruction and student progress; and improved understandings about how families and teachers might construct experiences that best facilitate children's literary meaning making, through talk, play, and other modalities; improved digital tools that elicit young children's literary meaning making; and improved research and pedagogy, especially for children who are English language learners and/or have disabilities.

• • •

Caitlin McMunn Dooley, Ph.D., is an Associate Professor at Georgia State University in the Department of Early Childhood Education. Her research focuses on the emergence of comprehension in young children. She can be reached at cdooley@gsu.edu.

References

Alexander, P. A. (2005/2006). The path to competence: A lifespan developmental perspective on reading. *Journal of Literacy Research*, *37*, 413-436.

American Academy of Pediatrics (2011). TV and Toddlers. Retrieved from http://www.aap.org/sections/media/toddlerstv.htm

Arizpe, E. (2001). 'Letting the story out': Visual encounters with Anthony Browne's The Tunnel. *Reading: Literacy and Language, 35*, 3, 115-119.

Beach, R. (2000). Critical issues: Reading and responding to literature at the level of activity. *Journal of Literacy Research, 32* (2), 237-251.doi: 10.1080/1086296000

Bearne, E. (2003). Rethinking literacy: Communication, representation an text. *Reading Literacy and Language, 37*(3), 98-103.

Bus, A. G., & Neuman, S. B. (2009). *Multimedia and literacy development: Improving Achievement for young learners*. NY: Routledge.

Cain, C., & Oakhill, J. (2007). *Children's comprehension problems in oral and written language: A cognitive perspective*. NY: Guilford.

Clay, M. M. (1991). *Change over time in children's literacy development*. Auckland, New Zealand: Heinemann.

Dooley, C. M. (2010). Young children's approaches to books: The emergence of comprehension. *The Reading Teacher, 64*, 2, 120-130. doi:10.1598/RT.64.2.4

Dooley, C. M., Martinez, M., & Roser, N. L. (in press). Young children's literary meaning making. In J. Larson and J. Marsh (Eds.) *Handbook of Early Childhood Literacy*. London: Sage.

Dooley, C. M. & Matthews, M. W. (2009). Emergent comprehension: Understanding comprehension development among young literacy learners. *Journal of Early Childhood Literacy, 9*, 269-294. doi:10.1177/1468798409345110

Dooley, C. M., Matthews, M. W., Matthews, L., & Champion, R. N. (2009). Emergent comprehension: The role of intentions during early print experiences. *National Reading Conference Yearbook*, *58*, 261-276.

Duke, N. K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly, 35,* 202–224.

Epstein, S. & Phillips, J. (2009). Storytelling skills of children with specific language impairment. *Child Language Teaching and Therapy, 25*, 285-300.

Evans, J. (Ed.) (2009). *Talking beyond the page: Reading and responding to picturebooks*. Abingdon: Routledge.

Flewitt, R., Nind, M., & Payler, J. (2009). 'If she's left with books she'll just eat them': Considering inclusive multimodal literacy practices. *Journal of Early Childhood Literacy, 9*, 211-233. doi:10.1177/1468798409105587

Galda, L., & Cullinan, B. (2002). *Literature and the child* (5th ed.). Belmont, CA: Wadsworth/Thomson Learning.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10.

Kaefer, T., Pinkham, A., & Neuman, S. B. (2010). Taxonomic organization scaffolds young children's learning from storybooks: A design experiment. *Society for Research on Educational Effectiveness*, 10pp.

Kendou, P., Bohn-Getter, C., White, M. J., & van den Broek, P. (2008). Children's inference generation across media. *Journal of Research in Reading, 31*, 259-272.

Kendou, P., Lynch, J. S., van den Broek, P., Espin, C. A., White, M. J., & Kremer, K. E. (2005). Developing successful readers: Building early comprehension skills through television viewing and listening. *Early Childhood Education Journal, 33,* 91-98. Doi: 10.1007/s10643-005-0030-6

Kendou, P., Savage, R., & van den Broek, P. (2009). *Revisiting the simple view of reading, 79*, 353-370. Doi:10.1348/978185408X369020

Korat, O., & Shamir, A. (2006). The educational electronic book as a tool for supporting children's emergent literacy in low versus middle SES groups. *Computers & Education, 50*, 110-124. Doi:10.1016/j.compedu.2006.04.002

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. London: Routletdge.

Labbo, L. D., & Kuhn, M. R. (2000). Weaving chains of affect and cognition: A young child's understanding of CD-ROM talking books. *Journal of Literacy Research, 32*, 2, 187-210. Doi:10.1080/10862960009548073

Landry, S. H., & Smith, K. E. (2006). The influence of parenting on emergent literacy skills. In D. K. Dickinson & S. B. Neuman (Eds.) *Handbook of early literacy research* (Vol. 2) (pp. 135-148). New York: Guilford Press.

Lankshear, C., & Knobel, M. (2003). New technologies in early childhood education literacy research: A review of research. *Journal of Early Childhood Literacy, 3*, 59-82.doi:10.1177/14687984030031003

Larson, L. (2010). Digital readers: The next chapter in e-book reading and response. *Reading Teacher, 64*, 1, 15-22.doi:10.1598/RT.64.1.2

Lonigan, C. J., & Shannahan, T. (2010). Developing early literacy skills: Things we know and things we don't know. *Educational Researcher, 39*, 340-346. doi:10.31102/0013189X10369832

Lynch, J. S., van den Broek, P., Kremer, K. E., Kendeou, P., White, M. J., & Lorch, E. P. (2008). The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology, 29*, 327-365.

Martinez, M., Roser, N. L., & Dooley, C. M. (2003). Young children's literary meaning making. N. Hall, J. Larson, and J. Marsh (Eds.), *Handbook of Early Childhood Literacy* (1st ed.) (pp. 222-234). London: Sage.

Martinez, M. & Teale, W. H. (1993). Teachers storybook reading style: A comparison of six teachers. *Research in the Teaching of English, 27*, 175-199.

Matthews, R., Dooley, C. M., & Czaplicki, K. (2011). Using parents' perceptions to gain insights into a young child's emergent literacy journey: A phenomenological case study. *Literacy Research Association Yearbook, 60*, 224-238.

McKeown, M. C., & Beck, I. L. (2006). Encouraging young children's language interactions with stories. In D. K. Dickinson & S. B. Neuman (Eds.) *Handbook of early literacy research* (Vol. 2) (pp. 281-294). New York: Guilford Press.

O'Neill, D. K., & Shultis, R. M. (2007). The emergence of the ability to track a character's mental perspective in narrative. *Developmental Psychology, 43*, 4, 1032-1037. doi:10.1037/0012-1649.43.4.1032

Palinscar, A. S., & Duke, N. K. (2004). The role of text and text-reader interactions in young children's reading development and achievement. *Elementary School Journal, 105*, 2, 183-197. doi:0013-5984/2004/10502-0004

Pantaleo, S. (2004). Young children interpret the metafictive in Anthony Browne's Voices in the Park. *Journal of Early Childhood Literacy, 4*, 211-233. doi: 10.1177/1468798404044516

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*, 184-202

Paris, S. & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly, 38*, 36-76.

Price, L. H., van Kleeck, A., & Huberty, C. J. (2009). Talk during book sharing between parents and preschool children: A comparison between storybook and expository book conditions. *Reading Research Quarterly, 44*, 171-194. Doi: dx.doi.org/10.1590/RRQ.44.2.4

RAND Reading Study Group (2002). *Reading for understanding: Toward an R & D program in reading comprehension*. Santa Monica, CA: RAND. Retrieved September 4, 2007, from www.rand.org

Reese, E., Leyva, D., Sparks, A., Grolnick, W. (2010). Maternal elaborative reminiscing increases low-income children's narrative skills relative to dialogic reading. *Early Education and Development, 21*(3). 318-342. Doi: 10.1080/10409289.2010.481552

Riggio, M. M., & Cassidy, K. W. (2009). Preschoolers' processing of false beliefs within the context of picture book reading. *Early education and development, 20* (6), 992-1015. doi:10.1080/10409280903375685

Rosenblatt, L. (1938/1965). *Literature as exploration*. New York: D. Applton-Century.

Roskos, K. A., Christie, J. F., Widman, S., & Holding, A. (2010). Three decades in: Priming for meta-analysis in play-literacy research. *Journal of Early Childood Literacy, 10*, 55-96. doi: 10.11771468798409357580

Serafini, F. (2010). Reading multimodal texts: Perceptual, structural, and ideological perspectives. *Children's Literature in Education, 41*, 85-104. doi:10.1007/10583-010-9100-5

Shamir, A. (2009). Processes and outcomes of joint activity with e-books for promoting kindergarteners' emergent literacy. *Media International, 46*, 1, 81-96. doi:10.1080/09523980902781295

Siegel, M. (2006). Rereading the signs: Multimodal transformations in the field of literacy education. *Language Arts, 84*, 1, 65-77.

Siegel, M., Kontovourki, S., Schmier, S., & Enriquez, G. (2008). Literacy in motion: A case study of a shape-shifting kindergartener. *Language Arts, 86*, 2, 89-98.

Sipe, L. R. (2000). The construction of literary understanding by first and second graders in oral response to picture storybook read-alouds. *Reading Research Quarterly, 35*, 2, 252-275.

Sipe, L. R. (2006). Young children's resistance to stories. *The Reading Teacher, 60,* 11, 6-13. doi:10.1598/RT.60.11

Sipe, L. R. (2008). *Storytime: Young children's literary understanding in the classroom*. NY: Teachers College Press.

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,* 97-122.

Stevens, R. J., Van Meter, P., &Warcholak, N. D. (2010). The effects of explicitly teaching story structure to primary grade children. *Journal of Literacy Research, 42*, 159-198.

Taylor, B. M., & Pearson, P. D. (2004). Research on learning to read—at school, at home, and in the community. *The Elementary School Journal, 105*, 2, 167-181. doi:0013-5984/2004/10502-0003

Tilstra, J., McMaster, K., van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading, 32*, 383-401. Doi: 10.1111/j.1467-9817.2009.01401.x

U.S. National Early Literacy Panel (2008). *Developing early literacy: A scientific synthesis of early literacy development and implications for intervention*. Washington, D. C.: National Institute for Literacy and National Center for Family Literacy. Retrieved online at: *lincs.ed.gov/publications/pdf/NELPReport09.pdf*

van den Broek, P. Kendeou, K. Kremer, J. S. Lynch, J. Butler, White, M. J. & Lorch, E. P. (2005). Assessment of Comprehension Abilities in Young Children in S. G. Paris & S. A. Stahl (Eds.) *Children's Reading Comprehension and Assessment* (pp. 131–160). Mahwah, NJ: Erlbaum.

Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice. *Australian Journal of Language and Literacy, 33*, 211-239.

Walsh, M. (2007). Reading digital texts. *Australian Journal of Language and Literacy, 30*, 40-53.

Wohlwend, K. E. (2008a) Kindergarten as nexus of practice: A mediated discourse analysis of reading, writing, play, and design in early literacy apprenticeship. *Reading Research Quarterly, 43,* 3, 332-334. doi:dx.doi.org/10.1598/RRQ.43.4.1

Wohlwend, K. (2008b). Play as a literacy of possibilities: Expanding meanings in practices, materials, and spaces. *Language Arts, 86*, 2, 127-636.

Wohlwend, K. E. (2009a). Early adopters: Playing new literacies and pretending new technologies in print-centric classrooms. *Journal of Early Childhood Literacy, 9*, 2, 117-140. doi:10.1177/1468798409105583

Wohlwend, K. E. (2009b). Damsels in discourse: Girls consuming and producing identity texts through Disney princess play. *Reading Research Quarterly, 44*, 1, 57-83. doi:dx.doi.org/10.1598/RRQ.44.1.3

Wohlwend, K. E. (2009c). Mediated discourse analysis: Researching young children's non-verbal interactions as social practice. *Journal of Early Childhood Literacy, 7*, 228-243. doi: 10.1177/1476718X09336950

Wohlwend, K. E. (2011). *Playing their way into literacies: Reading, writing, and belonging in the early childhood classroom*. NY: Teachers College Press.

Wyse & Styles (2007). Synthetic phonics and the teaching of reading: The debate surrounding England's Rose Report. *Literacy, 41*, 35-42.

Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., &Yapanel, U. (2009). Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics*, *124*, 342-349.

Children's literature cited:

Birney, B. G. (2006). *Friendship according to Humphrey*. NY: Puffin.

Browne, A. (1990/1997). *The tunnel.* Walker Books, Ltd.

Browne, A. (2001). *Voices in the park*. NY: DK Publishing.

International Electronic Journal of Elementary Education, 2011, 4(1), 185-195.

The effects of reading method on the comprehension performance of Saudi EFL students

Yousif ALSHUMAIMERI ∗
King Saud University, Saudi Arabia

Abstract

In this study I investigated the relative effects of different reading methods on the comprehension performance of Saudi EFL 10th grade male students. The scores of participants who read three comparable passages in three ways (oral, silent and subvocalizing) were compared. Results revealed a significant difference between oral reading and subvocalization, and between oral reading and silent reading. Oral reading had the greatest effect on comprehension performance among the three reading methods examined. All groups reported that oral reading was the most preferred reading method with the majority of respondents feeling the style best supported comprehension. Feedback suggested that oral reading was preferred specifically because it helps in memorizing words and texts, concentration, and practicing and pronouncing words for real world encounters. It is recommended that second language teachers and students use all available reading methods in order to identify which method best serves their study objectives.

Keywords: Oral reading, silent reading, subvocalization, reading comprehension, methods of reading, EFL, Saudi Arabia

Introduction

Reading ability has always been viewed as critical to academic success (Bernhardt, 1991; Carrell, 1991; Grabe & Stoller, 2002; Urquhart & Weir, 1998). Researchers investigating reading have attempted to look for components that affect reading performance as well as reading behaviors, such as oral reading, that distinguish proficient from less-proficient readers. Oral reading is often viewed as a dated methodology and discouraged by EFL/ESL teachers (Amer, 1997). While some researchers hold the opinion that oral reading is a way of wasting class time (Hill & Dobbyn, 1979), other scholars (Cho & Choi, 2008; Gibson, 2008; Rennie, 2000; Reutzel, Hollingsworth, & Eldredge, 1994; White, 1982) point to potential benefits that can be gained from various oral reading techniques that allow for oral proofreading, pronunciation practice, and conversational fluency.

∗ Yousif Alshumaimeri,College of Education, King Saud University, Riyadh, Saudi Arabia.
E-mail: yousif@ksu.edu.sa

ISSN:1307-9298

www.iejee.com

For decades, investigators have emphasized the importance of oral reading to children in first language teaching situations not only as a means of encouraging children to read, but also of improving their reading comprehension (Alshumaimeri, 2005; Grabe, 1991; Jackson & Coltheart, 2001; Juel & Holmes, 1981; McCallum, Sharp, Bell, & George, 2004; Prior & Welling, 2001; Rowell, 1976). According to Al-Qurashi, Watson, Hafseth, Hickman, & Pond (1995), in second language learning situations oral reading is the best way to teach pronunciation and word recognition during the early stages of second/foreign language acquisition, but reading comprehension is better strengthened by reading silently. Reading silently has traditionally been viewed as the only way to train pupils to read on their own (Al-Qurashi et al., 1995). The underlying principle governing this viewpoint is that reading is normally a solitary activity best done in total silence without interruption for best concentration (Al-Qurashi et al., 1995). While researchers continue to explore the effectiveness of oral reading on both language acquisition and comprehension, many questions remain unanswered.

Research on first language learning indicates that people often comprehend better after reading silently (Bernhardt, 1983; Leinhardt, Zigmond, & Cooley, 1981; Wilkinson & Anderson, 1995). However, other studies (Teng, 2009) suggest comprehension scores do not differ significantly between silent and oral reading. Further research on the relationship between reading methods and reading comprehension is needed in order to enhance EFL teaching methodologies and to improve learning outcomes. This research furthers understanding of the relationship between reading method and comprehension. As such, results will benefit educational institutions and the EFL researchers, teachers and students that support them.

Oral vs. Silent Reading Methods

Reading is a crucial skill in learning and communication. Current trends in education consider reading lessons to be an important early step in the development of mental and linguistic abilities. Reading methods include reading silently, reading using subvocalization (forming the sounds of the words while reading silently), and reading orally to oneself.

Reading silently means reading without labial movements or the vibration of vocal cords. This method implies that graphic forms are visually perceived and then transformed into meanings and ideas without passing through the vocal stage. Silent reading is usually seen as natural reading behavior and for decades has been associated with the idea of reading for comprehension.

As reviewed by Rennie (2000), academic work on reading pedagogy in the first half of the 20th century described the advantages, disadvantages and processes associated with both oral and silent reading (Chall, 1967; Russell, 1949). Although Russell (1949) found that in some places there was a system of reading called 'non-oral' which did not include oral reading instruction at any point in a child's reading development, most scholars agreed by the mid-twentieth century that both oral and silent reading activities were necessary for effective reading instruction.

Although the importance of oral reading to children learning a native language is widely accepted, the effectiveness of oral reading in second language classrooms continues to be debated. In her study of oral reading practices in the classroom, Gibson (2008) found that teachers and learners were using oral reading in a variety of ways. The primary reasons for using the method were for practicing pronunciation and intonation. Other reasons included for speaking practice, making graphemic-phonemic connections, diagnosing pronunciation problems, improving fluency and practicing reading skills. In the case of second language learning, Gibson (2008) also found that 82% of autonomous learners read orally to themselves as part of private study. Asian learners, in particular, commented that oral reading was especially important to them for practicing pronunciation.

Oral Reading and Comprehension

Hannon and Daneman (2001) proposed four primary processes in reading comprehension: accessing relevant knowledge from long-term memory, integrating accessed knowledge with information from the text, making inferences based on information in the text, and recalling newly learned text material. In schema theory, a predominant theory of reading comprehension, reading comprehension is viewed as the process of interpreting new information and assimilating this information into memory structures (Anderson & Pearson, 1984; Teng, 2009).

As suggested by Teng (2009), differences in native languages can affect second language (L2) oral reading for EFL learners. Reading in a second language requires more cognitive capacity for word identification than reading in one's native language (L1). Slower readers must employ greater cognitive resources than good readers in order to accomplish word recognition. Taguchi and Gorsuch (2002) found that while L1 readers tend to focus more on content words, L2 readers focus equally on content words and grammatical function words. As such, the limited cognitive capacity L2 readers allocate to word recognition tasks may impair their comprehension.

In her study of EFL students, Amer (1997) states that oral reading by the teacher helps readers discover units of meaning that arise from multi-word phrases rather than meaning that is derived from individual words. Oral reading also helps readers to see text as a whole with various levels of meaning rather than as a dissectible passage of graphic cues. Amer suggests that, with appropriate practice, students will gradually begin to realize that a higher level of comprehension can be achieved by reading larger meaningful units of texts. Oral reading performed by the teachers can additionally reinforce correct understanding of punctuation and intonation further strengthening student comprehension.

In researching the relevance of oral reading fluency to reading comprehension, Saiegh-Haddad (2003) conducted a study with 22 Arabic and 28 Hebrew native speakers, 19-25 years old, enrolled in intermediate EFL courses. By analyzing participant's oral reading skill with two texts, one in the participant's native language and one in English, the researchers aimed to determine if there was a difference in the relationship between oral reading skill and reading comprehension. Although there was no relationship found between oral reading fluency and reading comprehension in either Arabic or Hebrew reading (Saiegh-Haddad, 2003), in English, participants with oral reading fluency were found to have better reading comprehension.

Possible explanations for the above finding can be found in a study conducted by Miller and Smith (1985). Conducting a study on comprehension after reading orally and silently, Miller and Smith (1985) tested 94 second to fifth graders who read either at a low level, medium level, or high level. The results suggest that poor readers are better at comprehending when reading orally as compared to reading silently, and are more adept at answering inferential questions than they are at answering literal questions (Miller & Smith, 1985). Average readers in Miller and Smith's (1985) study read silently more proficiently than poor readers and were able to answer inferential and literal questions equally well. Good readers were found to be proficient at both oral and silent reading and best able to answer literal questions (Miller & Smith, 1985). The results of Miller and Smith's (1985) study suggest not only that literal comprehension is the best indicator of reading competence, but also that poor readers do benefit from the use of oral reading in the classroom.

In Taiwan, Teng (2009) studied the relationship between reading comprehension and reading methods and learning styles of EFL 12th grade male students. Teng (2009) found that most students can benefit from both silent and oral reading activities. Being that some students in the study benefited more from oral reading than others, Teng suggests that EFL teachers be more flexible in selecting various reading methods for use in the classroom setting. Teachers could support a mixture of oral and silent reading assignments that would allow students to engage in their preferred style.

Second language readers often read slowly and have under-developed oral production when compared to native speakers. Oral reading practice was found by Taguchi and Gorsuch (2002) to be more effective than other reading methods at increasing reading speed and comprehension among beginning L2 readers. However, Taguchi and Gorsuch (2002) were doubtful that oral reading of passages can be effective for older L2 readers as they read orally less often and may be less comfortable with the method.

Oral Reading, Culture and Environment

The viewpoint that oral reading has limited benefit to learners does not take into account the differing social and cultural backgrounds of students. As mentioned previously, there is an enduring opinion that readers who read silently comprehend the most because they both read and think. However, this view implies that students are incapable of thinking while reading orally. It might be more accurate to say that a student's attenuation to the social environment in which she is reading would have a greater bearing on her ability to concentrate than her inability to do both at once.

Alshumaimeri (2005) argues that oral reading is not necessarily a faulty reading method as suggested by Nuttall (1996, but, rather, is an effective aid to comprehension. In a study conducted by Alshumaimeri, oral reading was found to be used not only for decoding and relating written symbols to sounds, but also for comprehension. The criteria for effective reading comprehension included familiarity or comfortableness with the reading method, which aided the reader's speed of comprehension. Some informants in Alshumaimeri's study stated they would read orally when they were studying, which requires concentration, memorization, and comprehension, and would read silently when they read for enjoyment.

Furthering the body of work on L2 reading methods, this study investigates the effects of different reading methods on L2 student reading comprehension. The research questions are as follows: do different reading methods affect the comprehension of Saudi students; which reading methods affect reading comprehension; and which reading methods do Saudi students prefer and why?

Methods

Research Design

This research employs a classroom-based, quasi-experimental design in order to examine the effects of different reading methods on the comprehension performance of Saudi students. In educational research, a quasi-experiment is more commonly used due to fixed school schedules and logistical problems (Cohen, Manion, & Morrison, 2007). The different reading methods studied were oral, subvocalization, and silent reading. Comprehension performance was determined from the students' comprehension scores on multiple-choice tests. In order to minimize the effects of repetition, three different passages were selected from McCall-Crabbs Standard Test Lessons in Reading, Book D (1979).

Each group read each passage using one of the three reading methods (oral, subvocalization, and silent). The study participants always read passage 1 first, and then passages 2 and 3. However, in order to counterbalance the design of the study, the order of the reading method was rotated. For example, Group 1 read passage 1 orally, passage 2 using subvocalization, and passage 3 silently, while Group 2 read passage 1 using subvocalization, passage 2 silently, and passage 3 orally. In this way, the effect of passage difficulty or type of passage reading was minimized with regard to measuring the reading comprehension performance of the study students.

Each group was located in a different room during testing. The noise level during oral reading was not perceived as a distraction to comprehension as students read softly and the testing rooms were large. After reading each passage and taking the reading test for that passage, the students were asked to fill out a feedback slip asking about their preferred reading method and the reason behind their preference.

Participants

Participants in the study were 145 Saudi male students with an average age of 16 years, in the first year of the secondary stage (10th grade) in a secondary school in Riyadh, Saudi Arabia. Like most Saudi students they had studied English for four years, since the 6th grade in elementary school. Participating students were expected to be fairly representative of the target population of Saudi learners in terms of ability, interest, and age. However, one should acknowledge the limitation of drawing students from one school in Riyadh. The participating students were distributed by the school management into four classes with the intention that each class should be a balanced mixed-ability class. The students' level of language proficiency was considered to be A1 level of the European framework.

Passages

Three expository passages were selected from McCall-Crabbs Standard Test Lessons in Reading, Book D (1979). Each passage was followed by five questions, posed in English, with four-option multiple choice answers. Multiple choice questions are perhaps the most commonly used format in standardized reading comprehension tests. The procedure's advantages lie in the simplicity of its scoring (Koda, 2005). Passage one, A School Charity Day, contains 141 words and describes a fundraising bazaar held at a children's school. The second passage, The Best Way to Lose Weight, contains 139 words and provides advice on how to lose weight. The third passage, A Carpenter Story, contains 108 words and describes how one person became a carpenter. A reliability analysis was computed for each test using test/retest method (Cohen, Manion, & Morrison, 2007). The reliability results were (Pearson coefficient) 0.7462 for the first passage, 0.6715 for the second passage, and 0.6605 for the third passage. Reliability was deemed sufficient given that the test only contained five items.

Feedback Slips

The feedback slip was a small piece of paper that was given to each student after completing each reading test (three feedback slips were collected per participant). It included three questions that asked students to write down (in L1) if the reading method they used supported their comprehension, to rank which reading methods they generally prefer, and to explain their choices. The purpose of the feedback slips was to help in understanding the effects of the different reading methods and to know which reading methods students prefer in everyday life. The number of responses collected was 227 out of 435 feedback slips distributed with a return rate of 52.2%. The low rate of return is believed to be because the slips were distributed after each test. Some students returned the feedback slips blank because they had answered the question on the first slip and did not change their views. The slips were distributed after each test in order to provide equal opportunity for students to reflect on each reading method.

Procedure

The research was conducted on a regular school day during the extra-curricular activity time (the last two periods of one day per week). The available time for testing was 110 minutes. Each reading test was allocated 20 minutes followed by 5-7 minutes for filling out the feedback slips. The students were randomly assigned to their group. As described above, each group read a passage using each of the three different reading methods. All students were told to read the reading instructions carefully and to ask for clarification if needed. There were three teachers, one for each group, who helped administer the tests and explained the procedure clearly. Needed materials were prepared beforehand and placed in envelopes according to the study design. The researcher supervised the administration by moving from one room to another to check that the procedures were followed according to plan and to answer any questions.

The data collected consisted of the comprehension scores obtained from the five multiple-choice questions designed for each of the three passages as well as the data collected from the feedback slips. The data analysis was conducted in accordance with the research questions, all of which were concerned with comprehension performance as measured by the scores from the multiple-choice questions, the dependent variable. A one-way analysis of variance (ANOVA) was used to test the differences between the reading methods and a post-hoc analysis using the Scheffe test was conducted to locate the source of differences. Then, two-way ANOVA was used to test the differences between the groups with different reading methods. The study results are reported below.

Results

The results obtained are presented in accordance with the research questions, beginning with the first research question. In order to answer the first research question (Do different reading methods affect the comprehension of Saudi students?) a one-way analysis of variance (ANOVA) test was conducted with the post-hoc Scheffe test. There was a significant difference at level 0.01 between the Saudi students in comprehension performance according to the reading method. A significant difference was found between oral reading and subvocalization (mean difference 1.92, $p < 0.01$), and between oral reading and silent reading (mean difference 2.32, $p < 0.01$). The largest mean occurred for oral reading (9.65), which had the greatest effect on comprehension performance among the three reading methods included in the study.

To answer the second research question, the data obtained from the feedback slips show the students' responses to the question (Does this reading method assist you in understanding this passage?). The results show that 57% of the students thought that oral reading helped them better comprehend the passage; whereas 26.2% and 17.9%, respectively, thought silent reading and subvocalization helped them understand the passage. The reading method that had the greatest positive effect on comprehension was oral reading with a mean value 9.65. Subvocalization and silent reading had mean values of 7.72 and 7.33, respectively. These results indicate that oral reading helped students better understand passages.

To answer the third research question (Which reading methods do Saudi students prefer and why?), the results obtained from the feedback slips show the ranked order of the preferred reading style of each group as well as an explanation of their choice of order. Of all groups, 50.57% of students reported that oral reading was the most preferred reading method. Subvocalization was ranked second with 22.76%, whereas silent reading was third with 14.02%.

Data obtained from the feedback slips is summarized in Tables 1a, 1b, and 1c and indicate the reason the students preferred each method of reading. The rate of return (52.2%) of the feedback slips could indicate that the reason for preferring a reading method is static and that the learners felt they did not need to provide the same feedback after each passage. The results are presented according to each reading method.

Table 1a: *The Students' Responses for Reasons for Preferring Oral Reading*

No.	Students justification	Frequency	Percentage
1	It helps in memorizing and remembering new words	26	20.63%
2	I use this way for studying as it helps me understand and memorize the text	25	19.84%
3	It makes me concentrate more and understand the text	20	15.9%
4	In reading aloud it helps me understand more as I use three senses (sight, hearing, and speech)	15	11.9%
5	It helps in pronunciation practice and pronouncing the words better	13	10.32%
6	It makes the words more familiar and helps memorizing them and using them in conversations with others	12	9.52%

Table 1a: *The Students' Responses for Reasons for Preferring Oral Reading*

No.	Students justification	Frequency	Percentage
7	I read faster and understand more in reading aloud	8	6.34%
8	It helps in conversation and practice talking in a foreign language	7	5.55%
Total		126	100%

Table 1b: *The Students' Responses for Reasons for Preferring Silent Reading*

No.	Students justification	Frequency	Percentage
1	I read silently for leisure not for study	11	29.73%
2	It helps me understand and concentrate more	10	27.03%
3	I read faster and understand more	6	16.22%
4	I don't like annoying other people when I read aloud	5	13.51%
5	I feel more relaxed when I read silently	5	13.51%
Total		37	100%

Table 1c: *The Students' Responses for Reasons for Preferring Subvocalization*

No.	Students justification	Frequency	Percentage
1	It makes me concentrate more and understand the text	15	23.43%
2	I use this way for studying	15	23.43%
3	I do not annoy other people and concentrate than reading silently	11	17.2%
4	It helps in memorizing and remembering new words	10	15.63%
5	It helps in pronunciation practice and pronouncing the words better	7	10.94%
6	I read faster than reading aloud and keep my concentration	6	9.37%
Total		64	100%

Table 1a shows that just over 40% of the respondents preferred oral reading because it was perceived as aiding in memorizing and remembering new words. Using oral reading as a means to improve conversational English or improve pronunciation was not a common justification among respondents. The usefulness of oral reading in improving conversational English was the least cited justification with only 5.5% of respondents choosing it as their primary reason for preferring oral reading.

As seen in Table 1b, nearly 30% of respondents indicated that they read silently for leisure, but not for study in justifying their preference for silent reading. Such a result indicates that students associate silent reading with leisure reading and oral reading with language studies. Many students, 27%, also indicated that understanding and concentration were heightened when reading silently. Anxiety about annoying others and greater relaxation while reading silently were less cited reasons for preferring silent reading (13.5% and 13.5%, respectively).

Respondents with a preference for subvocalization cited the justifications of enhanced concentration and preferred method of studying (23.4% and 23.4%, respectively) as seen in Table 1c. Such results indicate subvocalization is a study habit perceived as enhancing students' ability to concentrate on text while studying. As in the results presented in Table 1a summarizing the justifications for preferring oral reading, pronunciation was not an often cited reason for employing subvocalization as a reading method (10.9%).

A one-way ANOVA with post-hoc Scheffe test was used to explore the relationship between the preferred reading method and the comprehension performance of the students. A significant relationship was found between the preferred reading method and the students' comprehension performance, (F = 5.919, Sig. = .001). A significant difference was found in favor of the students who prefer oral reading over reading using subvocalization or silent reading (mean value 8.96). The results suggest that there is a relationship between the selected reading method and the comprehension performance of the Saudi students.

To evaluate if there is a significant difference between groups with regard to comprehension performance regardless of reading method, a two-way ANOVA was conducted. The results indicate three things: the differences between the groups, regardless of reading methods, were not significant; the differences based on reading methods were significant; and the interaction effects between groups and reading methods were significant.

Discussion

In summary, the study of the relative effects of different reading methods on the comprehension performance of Saudi EFL 10th grade students shows that there is a significant difference between the Saudi students in comprehension performance according to reading method. These results support the literature suggesting oral reading can be a beneficial reading method when used in the L2 classroom (Cho & Choi 2008; Gibson, 2008; Rennie, 2000; Reutzel, Hollingsworth, & Eldredge, 1994; White, 1982). If reading comprehension can be defined as the process of interpreting new information and assimilating this information into memory structures as schema theory suggests, this study indicates that oral reading aids comprehension by improving students' ability to concentrate and memorize new words.

In this study there was a significant difference between oral reading and subvocalization in regard to comprehension and between oral reading and silent reading. Oral reading had the greatest positive effect on comprehension performance among the three reading methods included in the study. As found in the literature, students exposed to oral reading techniques in L2 classrooms report improved comprehension of reading material (Amer, 1997; Saiegh-Haddad, 2003; Warwick & Mangubhai, 1983). Although it should be reiterated that reading ability was not measured among the participants in this study, all participants were in the same grade and differences in reading ability would presumably have an insignificant impact on the study results.

Despite the negative opinion some scholars hold toward the use of oral reading in the language classroom (Hill & Dobbyn, 1979), this study demonstrates that oral reading can in fact assist some students in acquiring proficient language comprehension. Oral reading, although often described as an effective method of learning for recognizing and pronouncing words with ease and fluency, was more often cited by the participants in this study as a means of strengthening memorization of new words and effective concentration as well as comprehension. The results of this study are in relation to student performance on a comprehension test and not in relation to language instruction or leisure reading.

Although many students indicated that understanding and concentration are heightened when reading silently, one third of respondents indicated that they read silently for leisure, but not for study. Such a result indicates students associate silent reading with leisure reading and oral reading with language studies. This preference for silent leisure reading supports Nuttall's (1996) opinion that oral reading is uncommon outside the classroom.

With regard to student preference for a particular reading style, all groups reported that oral reading was the most preferred reading method with subvocalization ranked second and silent reading third. These results suggest that most of the students hold the opinion that oral reading is an effective method for understanding the passages. In addition, participants indicated that oral reading was the preferred reading method not only because it helps in memorization and concentration, but it is also helpful for practicing and pronouncing words for real world encounters. This clear preference for oral reading for study purposes is partially due to traditional teaching methods, such as rote learning, that require learners to memorize information.

Respondents with a preference for subvocalization cited the justifications of enhanced concentration and preferred method of studying. Such results indicate subvocalization is a study habit that is perceived as enhancing students' ability to concentrate on text while studying. However, Saudi students do not seem to support Nuttall's (1996) opinion that subvocalization is an ineffective reading method. Participants in this study preferred subvocalization above silent reading as a reading method.

If Gibson (2008) and Amer (1997) are correct in assuming that oral reading can be made a more effective learning device with greater systemization in the curriculum, the methods by which Saudi teachers encourage and use oral reading in the classroom should be further researched. It is possible that the Saudi scholastic environment provides an ideal setting for allowing oral reading methods to support significant gains in reading comprehension. This educational setting, coupled with a cultural appreciation for oral religious traditions and memorization through oral recitation, could foster strong tendencies among students to associate oral reading with concentration and memorization.

As reported by Alshuamimeri (2005), one interviewee reported feeling as if oral reading allowed for better understanding and concentration. Further, the respondent suggested that Arabic literature in particular is better appreciated and analyzed when read aloud. Additionally, oral reading makes a strong impression because the reader hears as well as sees what is being read and the sense of hearing is effective in supporting comprehension. Albar (1996) stated the capacity to learn a language is dependent on normal hearing more than any other trait suggesting that someone who reads aloud is more likely to understand what he is reading than someone who reads silently.

Conclusion

In investigating the relative effects of different reading methods on the comprehension performance of Saudi EFL 10th grade students, the results showed that oral reading had the greatest positive effect on the comprehension performance of the study sample. This study also found that oral reading was the most preferred reading method. Oral reading was perceived by students to aid in memorization and concentration. EFL teachers can take from this finding that despite the relative inconsistencies in academic findings regarding the effect of reading method on comprehension, some students do in fact find oral reading to be beneficial in the L2 classroom. Although such findings may be culturally or individually specific, greater flexibility in the design of second language teaching methodologies is warranted pending greater research on the subject. Additionally, L2 students should use all available reading methods in order to identify which method best serves their study objectives.

Reading ability is acquired through practice, not through educational settings or teaching methods. When viewed as a continuum with beginners at one end and fluent readers at the other end, a student's growing capacity is defined by his or her ability to rapidly understand and comprehend new lexicon and context. Readers may find that while comprehension is not necessarily bolstered by practicing oral reading methods, memorization and concentration may be enhanced by employing oral reading techniques, either in the classroom or during private study. Unfortunately, the Arabic library is lacking material for children. Making reading materials available from an early age should be prioritized in order to support a well rounded adult ability to comprehend written language. To further the findings of this study, additional research is needed on L2 learners of different ages and gender.

• • •

Yousif Alshumaimeri is an associate professor at the College of Education, King Saud University, Saudi Arabia. He is head of the TESOL team at the department of Curriculum and Instruction and coordinator of the MA TESOL program. His main research interests are in teaching reading skills and the use of pedagogic tasks.

References

Albar, M. A. (1996). *Human Development*. Saudi publishing & distributing house, Jeddah, Saudi Arabia.

Al-Qurashi, K., Watson, M., Hafseth, J., Hickman, D,. & Pond, R. (1995). *'English For Saudi Arabia: first year secondary, Teacher's book'* .Ministry Of Education Press, Riyad, Saudi Arabia.

Alshumaimeri, Y. (2005). Is reading aloud a faulty habit. Paper presented at the second international conference: Language, culture and literature, Minia University, Egypt.

Amer, A. (1997). The effect of the teacher's reading aloud on the reading comprehension of EFL students. *ELT Journal,* 51(1), 43-47.

Anderson, R.C. & Pearson, P.D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P.D. Pearson (Ed.), *Handbook of Reading Research* (pp. 255-292). New York: Longman.

Bernhardt (1983). Three approaches to reading comprehension in intermediate German. *The Modern Language Journal,* 67, 111-115.

Bernhardt, E. B. (1991). *Reading development in a second language: Theoretical, empirical, and classroom perspectives.* Norwood, NJ: Ablex.

Carrell, P. L. (1991). Second language reading: Reading ability or language proficiency? *Applied Linguistics,* 12(2), 159-79.

Chall, J. (1967). *Learning to Read: The Great Debate*, McGraw-Hill, New York.

Cho, K.S. & Choi, D.S. (2008). Are Read-Alouds and Free Reading "Natural Partners"? *Busan National University of Education Research Institute,* 36(5), 69-73.

Cohen, L., Manion, L., & Morrison, K. (2007). Research methods in education (Sixth ed.). US: New York: Routledge.

Gibson, S. (2008). Reading aloud: a useful learning tool? *ELT Journal,* 62(1), 29-36.

Grabe, W. (1991). Current Developments in Second Language Reading Research. *TESOL Quarterly*, 25(3), 375-406.

Grabe, W. & Stoller, F. L. (2002). *Teaching and researching reading.* London: Longman.

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(1),* 103-128.

Hill, L. and Dobbyn, M. (1979). *A Teacher Training Course for Teachers of EFL*. London, UK: Cassell.

Jackson, N. & Coltheart, M. (2001). *Routes to reading success and failure*. Hove, UK: Psychology Press.

Juel, C. & Holmes, B. (1981). Oral and silent reading of sentences. *Reading Research Quarterly,* 16(4), 545-568.

Koda, K. (2005). *Insights into second language Reading: A cross-linguistics approach.* Cambridge, UK: Cambridge University Press.

Leinhardt, G., Zigmond, N., & Cooley, W. W. (1981). Reading instruction and its effects. *American Educational Research Journal,* 18*,* 343-361.

McCall, W. & Crabbs, L. (1979). *Standard Test Lessons in Reading*, Book D. New York, US: Teachers College, Columbia University,

McCallum, R. S., Sharp, S., Bell, S. M., & George, T. (2004). Silent versus oral reading
comprehension and efficiency. *Psychology in the Schools,* 41(2), 241-246.

Miller, S. D., & Smith, D. E. P. (1985). Differences in literal and inferential comprehension after reading orally and silently. Journal of Educational Psychology, 77(3), 341-348.

Nuttall, C. (1996). *Teaching Reading Skills in a Foreign Language.* UK: Heinemann.

Prior, S. M., & Welling, K. A. (2001). "Read in your head": A Vygotskian analysis of the transition from oral to silent reading. *Reading Psychology,* 22(1), 1-15.

Rennie, J. (2000). Teaching reading: Oral reading practices as a pedagogical tool in the primary school. *The Australian Journal of Language and Literacy*, 23(3), 197-211.

Reutzel, D. R., Hollingsworth, P. M. & Eldredge, J. L. (1994). Oral Reading Instruction: The Impact on Student Reading Development. *Reading Research Quarterly*, 29(1), 40-62.

Rowell, E.H. (1976). Do elementary students read better orally or silently? *The Reading Teacher Journal,* 29(4), 367-370.

Russell, D. H. (1949). *Children Learn to Read.* Boston: Ginn and Company.

Saiegh-Haddad, E. (2003). Bilingual oral reading fluency and reading comprehension: The case of Arabic/Hebrew (LI) -English (L2) readers. *Reading and Writing: An Interdisciplinary Journal,* 16(8)*,* 717-736.

Taguchi, E., & Gorsuch, G. J. (2002). Transfer effects of repeated EFL reading on reading new passages: A preliminary investigation. *Reading in a Foreign Language,* 14(1), 43-65.

Teng, Y. (2009). The Relationship of reading methods and learning styles to Taiwanese 12th Grade male students' reading comprehension in English. A Dissertation presented to The Faculty of the School of Education International and Multicultural Education Department. The University of San Francisco.

Urquhart, S. & Weir, C. (1998). *Reading in a second language: Process, product and practice.* London: Longman.

Warwick, B. E. & Mangubhai, F. (1983). The impact of Reading on second language learning. *Reading Research Quarterly,* 19 (1), 53-67.

White (1982). *The English Teacher's Handbook.* Walton on Thames: Nelson.

Wilkinson, I. A. G., & Anderson, R. C. (1995). Sociocognitive processes in guided silent reading: A microanalysis of small-group lessons. *Reading Research Quarterly,* 30(4), 710-740.

www.iejee.com

This page is intentionally left blank

International Electronic Journal of Elementary Education, 2011, 4(1), 197-227.

Children's comprehension of informational text: Reading, engaging, and learning

Linda BAKER∗
University of Maryland, Baltimore County, United States

Mariam Jean DREHER
University of Maryland, College Park, United States

Angela Katenkamp SHIPLET
University of Maryland, Baltimore County, United States

Lisa Carter BEALL
University of Maryland, Baltimore County, United States

Anita N. VOELKER
University of Maryland, College Park, United States

Adia J. GARRETT
University of Maryland, Baltimore County, United States

Heather R. SCHUGAR
University of Maryland, College park, United States

Maria FINGER-ELAM
University of Maryland, Baltimore County, United States

Abstract

The Reading, Engaging, and Learning project (REAL) investigated whether a classroom intervention that enhanced young children's experience with informational books would increase reading achievement and engagement. Participants attended schools serving low income neighborhoods with 86% African American enrollment. The longitudinal study spanned second through fourth grades. Treatment conditions were: (1) Text Infusion/Reading for Learning Instruction -- students were given greater access to informational books in their classroom libraries and in reading instruction; (2) Text Infusion Alone -- the same books were provided but teachers were not asked to alter their instruction; (3) Traditional Instruction -- students experienced business as usual in the classroom. Children were assessed each year on measures of reading and reading engagement, and classroom instructional practices were observed. On most measures, the informational text infusion intervention did not yield differential growth over time. However, the results inform efforts to increase children's facility with informational text in the early years in order to improve reading comprehension.

Keywords: Reading comprehension, informational text, reading instruction

* Linda Baker, Department of Psychology, UMBC 1000 Hilltop Circle Baltimore, MD 21250, USA. Phone: 410-455-2370. E-mail: baker@umbc.edu

ISSN:1307-9298

www.iejee.com

Introduction

The tremendous nationwide emphasis on early reading achievement that came about through No Child Left Beyond had as a primary focus the development of the basic skills of word recognition. The ability to read words fluently and automatically is critical to achievement, but it is not sufficient. Nor is the ability to comprehend stories sufficient. Although stories comprise a major proportion of the materials children encounter in early reading instruction (Duke, 2000; Hoffman et al., 1994; Jeong, Gaffney, & Choi, 2010, Moss, 2008; Moss & Newton, 2002; Ness, 2011), older students are expected to comprehend and learn from informational text, and they must continue to do so as they move through school and into the job market (Common Core State Standards (CCSS), 2010; Salinger, Kamil, Kapinus, & Afflerbach, 2005; White, Chen, & Forsyth, 2010; Venezky, 2000). Thus, facility with informational text is also critical to achievement, in school and beyond.

The Reading, Engaging, and Learning project (REAL) that is the focus of this article was designed to respond to the national goals of raising reading achievement and closing the achievement gap. The intervention study was funded by the Spencer Foundation in 2000, at a time when national experts had begun to recommend that children be given more exposure to informational text in the early years of schooling (e.g., Snow, Burns, & Griffin, 1998), but when support for this recommendation was mostly anecdotal and correlational (Campbell, Kapinus, & Beatty, 1995; Caswell & Duke, 1998). The REAL project was a multi-component intervention study that yielded information about a variety of facets of children's experiences with informational text, including comprehension of text as assessed on a standardized multiple-choice instrument and on a performance assessment, ability to use the features of informational text that enhance comprehension, reading motivation, out-of-school reading activity, instructional practices of the classroom teachers, preferences for reading, and gender differences.

The three-year longitudinal study, spanning grades two through four, yielded only limited evidence of an effect of the intervention to increase students' access to and comprehension of informational text (Baker & Dreher, 2005; Dreher & Baker, 2005). For that reason, we did not pursue publication of the REAL study at the time it concluded in 2005. However, this decision did not mean that we thought efforts to enhance children's comprehension of informational text were bound to be unsuccessful. Rather, we attributed the null effects primarily to school and teacher factors beyond our control. With hindsight, we have come to realize that the study provides much valuable information about children's understanding of informational text, independent of intervention effects. Accordingly, the purpose of this article is to describe the study in the context of the growing body of knowledge on informational text comprehension and use in the early elementary grades.

Traditionally, learning through reading has been delayed until children have learned how to read, using familiar topics in stories. But expert opinion has converged on the notion of integrating learning to read and reading to learn (e.g., CCSS, 2011; International Reading Association/National Council of Teachers of English, 1996; National Council for the Social Studies, 1994; National Council of Teachers of Mathematics, 1989; National Research Council, 1994; Snow et al., 1998; Salinger et al., 2005). Changes have begun to appear in the past decade, but the elementary school reading experience continues to be predominantly stories (Jeong, et al., 2010, Moss, 2008, Ness, 2011). Analyses of classroom reading materials suggest that up to 90% of what is read in the classroom consists of stories (Dreher, 2000), yet there is no compelling reason why this should be so, even in the primary grades.

Research shows that young children can and do appreciate and understand informational text (Cervetti, Bravo, Hiebert, Pearson, & Jaynes, 2009; Diakidoy, Stylianou, Karefillidou, &

Papageorgiou, 2005; Horowitz & Freeman, 1995; Pappas, 1993; Smolkin & Donovan, 2001). Furthermore, they can be taught strategies for learning from it (Aarnoutse & Schellings, 2003; Culatta, Hall-Kenyon, & Black, 2010; Duke & Carlisle, 2011; Hall, Sabey, & McLellan, 2005; O'Hara, 2007; Williams et al., 2005; Williams, Stafford, Lauer, Hall, & Pollini, 2009). The new Common Core State Standards (CCSS, 2010) specify that children should acquire competencies with informational text beginning in kindergarten. For example, with prompting and support, kindergartners should be able to ask and answer questions about key details in a text, ask and answer questions about unknown words in a text, and identify basic similarities in and differences between two texts on the same topic (e.g., in illustrations, descriptions, or procedures). The fact that informational text processing skills are to be part of the standard language arts curriculum is a powerful indicator of the importance of fostering children's comprehension of such text from the earliest years of formal schooling.

Indeed, early attention to informational text is needed if by fourth-grade children are to be "capable—independently and productively—of reading to learn" (Snow et al., 1998, p. 207). The National Assessment of Educational Progress (NAEP) now reflects this expectation (Foorman & Connor, 2011). Beginning with the 2009 reading assessment, the NAEP framework for fourth grade specifies a 50-50 proportion of total testing time devoted to reading informational vs. literary text (National Assessment Governing Board, 2008), compared to a 45-55 allocation in the 1992 to 2007 frameworks. In addition, for the first time, the 2009 NAEP reports scores separately by genre.

Performance on the NAEP, however, indicates the schools have far to go. The 2009 NAEP found two thirds of American fourth graders can only read at or below a basic level (33% below basic and 34% at basic), meaning they exhibit only partial mastery of reading skills (NCES, 2009). This figure has decreased only slightly since 1992, when the percentage at or below basic was 71%, indicating that the recent federal initiatives to improve early reading skills have met with limited success. In fact, evaluations of Reading First have revealed that benefits do not extend beyond decoding to reading comprehension (Gamse, Jacob, Horst, Boulay, & Unlu, 2008).

The 2009 NAEP results also document that certain subgroups of fourth graders continue to perform lower than the national average. For example, 52% of black and 51% of Hispanic children scored *below* a basic level in reading, whereas 22% of white and 20% of Asian/Pacific Island children did so. Furthermore, 71% of English language learners (ELL) scored below basic in comparison to 30% of non-ELLs, as did 49% of children eligible for free or reduced-cost meals, in comparison to 20% of those not eligible (http://nationsreportcard.gov/reading_2009/). Thus, overall fourth-grade reading performance needs improvement, but for poor and minority children the situation is particularly urgent.

To pursue our specific interest in comprehension of informational text, we accessed the NAEP data base and used their statistical tools to compare scale scores achieved on literary and informational text across selected demographic subgroups. The data for fourth graders on the 2009 assessment are provided in Table 1. Overall, students scored better on the literary passages than the informational. Note that the demographic disparities in performance are even greater on informational text than on literary text. For example, children eligible for free lunch scored 28 points lower than non-eligible children on informational text, compared to 25 points lower on literary text.

Students from high-poverty areas often lack exposure to the academic vocabulary needed for reading comprehension (Chall, Jacobs, & Baldwin, 1990; Neuman, 2006). Informational text is a major source of difficult, abstract, specialized, and technical words.

The narrow focus on word recognition so typical in early literacy instruction should be replaced by content-rich environments that include informational books, given their considerable potential for increasing children's background knowledge and conceptual development (Foorman & Connor, 2011; Neuman, 2001; 2006; 2010).

Focusing on both learning to read and reading for learning -- with early, balanced attention to both stories and informational text – may increase children's reading achievement (CCSS, 2010; Dreher & Voelker, 2004). Support for this suggestion comes from the NAEP, where it was found that on the 1992 assessment, fourth graders who reported reading not only stories, but also magazines and informational books had the highest achievement (Campbell et al., 1995).

Table 1. *Mean Scale Scores for Fourth Graders on the 2009 NAEP Reading Assessment by Type of Text and Selected Demographic Variables*

Demographic Variable	Subgroup	Literary Text	Informational Text
Full Sample	------	222	219
Income	Eligible for free lunch	208	203
	Not eligible	233	231
Language	ELL	190	185
	Non-ELL	225	222
Race	White	231	229
	Black	207	202
	Hispanic	208	202
	Asian/Pacific Isl.	236	234
Gender	Girls	226	222
	Boys	218	216

Note. Maximum scale score is 500. Differences as a function of income level, language status, and gender were statistically significant at $p < .001$ for each text type. For the Race variable, Asian/Pacific Islanders had the highest scores, followed by whites; Blacks and Hispanics did not differ significantly. Differences across text types were not analyzed due to limitations of the NAEP online software.

An important enabler of reading comprehension is reading engagement (Baker, Dreher, & Guthrie, 2000). Engaged readers read widely and frequently, and they seek opportunities to learn from reading. Their motivations include the beliefs, desires, and interests that lead them to choose to read. Yet reading motivation drops as children move through the grades (Chapman & Tunmer, 1997; McKenna, Ellsworth, & Kear, 1995; Wigfield et al., 1997). This decline occurs for all achievement levels but is most pronounced for children of lower reading ability. Effective use of more diverse material, including informational books, may help to counteract this drop in motivation to read by arousing children's interest and curiosity, which in turn should increase independent reading activity (Baker & Wigfield, 1999; Duke & Carlisle, 2011; Wigfield & Guthrie, 1997). Guthrie and Wigfield and their colleagues have provided compelling evidence that classroom interventions that increase reading engagement also increase reading comprehension (Guthrie, Wigfield, Barbosa et al., 2004). Furthermore, the more motivated the reader, the more growth in comprehension over time (Guthrie, et al., 2007).

A variety of frameworks have been used to categorize text types as fiction vs. nonfiction or literary vs. informational. The guidelines of Kletzien and Dreher (2004) were used in the REAL project. The informational text category includes narrative-informational, expository, and mixed. Narrative-informational text uses a story or narrative format to convey factual information. Expository texts do not include story elements such as characters, goals, and resolutions. Instead they might be characterized as reports, using text structures such as cause and effect, comparison and contrast, sequence, description, and problem and solution. They explain things about the natural and social world such as animals, places, and cultural groups. Mixed texts, also referred to as dual-purpose, blended, or hybrid texts, mix narrative and expository writing in the same texts. For example, in *The Popcorn Book* (dePaola, 1984), the story of brothers making popcorn is accompanied by encyclopedia-like facts about popcorn which one brother reads aloud.

The REAL project was based on the premise that increasing students' access to informational text would enhance their knowledge and comprehension. Accordingly, the central thrust of the project involved an infusion of informational books into classroom libraries. Classroom libraries are frequently limited in the resources they offer, and this is particularly true in schools serving low income neighborhoods (Neuman, 2006). The effective classroom library provides a variety of book genres including informational books, stories, poetry, references, and multimedia, and it offers a range of difficulty levels and caters to a range of interests (Dreher & Voelker, 2004). Given the opportunity to access diverse materials in the classroom library, student interest in reading should increase, the amount of time spent reading should increase, and gains in reading skills and strategies should be realized (Chambliss & McKillop, 2000). Of course, at the same time, teachers need to provide instruction designed to help children read and learn from diverse genres. Accordingly, one of the treatment conditions included professional development for teachers on reading for learning.

The project was a three-year longitudinal study, commencing in second grade and continuing through fourth grade. Second grade was selected as the starting point because children at this level have begun to master the skills of word recognition and are capable of comprehending other than the simplest of texts. Fourth grade was selected as the ending point because it is at this level that many children begin to struggle, as the demands shift more from learning to read to reading to learn. It is not our intention in this article to provide a comprehensive empirical report of the entire REAL project but rather to describe certain components of the project in greater depth than others, consistent with the theme of this special issue on reading comprehension. Students within classrooms were assigned to one of three treatment conditions. It was expected that students who experienced both text infusion and reading for learning instruction would exhibit the greatest gains in reading achievement and engagement; students who received text infusion alone would also benefit relative to those receiving traditional instruction, but not to the same extent as those in the combined treatment condition.

Overview of the Methods of the REAL Project

Treatment Conditions

Informational text infusion/Reading for learning (RFL) instruction (Text infusion/RFL Instruction). Classroom libraries were enhanced with informational books, and teachers participated in professional development sessions on reading for learning, as described below. Students were assessed on reading comprehension using both standardized and researcher-developed measures, motivation for reading, and reading activity each year of the project. Students kept logs of books read during independent reading time, and teachers kept logs of books read aloud to the students. Classrooms were visited regularly to observe instruction and collect logs.

Informational text infusion/Traditional instruction (Text infusion alone). To control for the possibility that the informational text infusion alone might be responsible for any positive outcomes (i.e., book flood effects, Ingham, 1982), classroom libraries were enhanced with the same informational books as in the RFL instructional condition, but teachers were not asked to modify their usual instructional practices. All of the same data were collected.

Traditional Instruction. Classroom libraries were not supplemented with informational text, and teachers were not asked to modify their usual instructional practices. The same assessments were used, but students and teachers did not keep logs of their reading as this in itself might influence outcomes. However, classroom observations were made to provide data on informational book availability and instruction relating to informational text. (Teachers received books of their choice at the end of the school year for use with the next year's students.)

Participants

The study was conducted within three demographically similar public elementary schools in a large suburban school system with a majority African American student population. The three schools that were selected also had a majority of children receiving free or reduced price lunch and a majority scoring below average on statewide assessments. The ethnicity distribution of participants at project outset was 86% African-American, 10% European-American, 2% Hispanic-American, and 1% Asian-American. During the first year of the project, 222 second-grade children participated. In subsequent years, the number of participants was affected by withdrawals, new enrollees, and redistricting. In Year 2, 195 third graders participated, and in Year 3, 209 fourth graders. Each school was randomly assigned to one of the three treatment conditions. Table 2 shows the distribution of classrooms and teachers across conditions and years, as well as the number of students remaining in the REAL project for all three years.

Table 2. *Number of Classes and Students in the Reading, Engaging, and Learning (REAL) Project by Treatment Condition*

Project Year	Text Infusion/RFL Instruction		Text Infusion Alone		Traditional Instruction	
	Classes	Students	Classes	Students	Classes	Students
1: Grade 2	2	60	3	77	4	85
2: Grade 3	3	62	4	79	3	69
3: Grade 4	2	51	4	69	3	89

Note. The number of students participating in the project across all three project years was 28 in Text Infusion/RFL, 51 in Text Infusion Alone, and 30 in Traditional Instruction.

Teacher Training

At the beginning of each school year, brief orientations were given to the teachers who were participating in the project at all three schools, followed by in-service sessions with teachers in the text infusion schools. For the Text Infusion Alone teachers, a single short session focused on procedures for completing the teacher and student reading logs. Text infusion/RFL Instruction teachers received in-service sessions on RFL instruction, one in the fall at the beginning of the intervention and one after winter vacation. In addition, we held informal discussions during our regular visits to the classrooms and via email. Key instructional components were selected on the basis of research evidence attesting to their promise: (1) balancing the use of informational books in daily read-alouds; (2) promoting diverse daily independent reading; and (3) teaching children strategies and text features for comprehending and using informational text.

At second-grade, the fall session included modeling and providing guidelines on how to use information books in read-alouds (e.g., teach children to use text access features, activate prior knowledge, highlight new vocabulary, engage children in discussion). We explained the Questioning the Author (QtA) approach to enhance children's comprehension (Beck McKeown, Hamilton, & Kucan, 1997). We also discussed examples of different types of informational text (narrative-informational, expository, and mixed) and the importance of instruction on expository text. In the second-grade winter session, we continued working with teachers to offer suggestions and address any concerns on the earlier topics. In addition, we introduced ways to use informational books to encourage expository writing.

At third and fourth grades, in-service sessions continued attention to the same topics but with additional foci. At third-grade, we modeled the use of a variety of I-charts (Hoffman, 1992) for organizing new information. We also provided additional techniques to support vocabulary development, and more ways to encourage expository writing such as paragraph frames (Lewis, Wray, & Rospigliosi, 1994). At fourth grade, we provided teachers with a strategy instruction plan, overviewed key comprehension strategies for use with informational text, reviewed common expository text structures, and modeled how to help children search for and monitor their hunt for information (Kletzien & Dreher, 2004). In all grades, teachers received handouts, and a notebook in which to keep them, including a summary of research support for each topic, specific guidelines, and examples.

Availability of Informational Text in Classrooms

Classroom libraries were inventoried each year, at the beginning before books were provided in the text infusion conditions and at the end. Books were counted, listed, and categorized by type. Classroom libraries varied greatly in size, but most were quite modest. As second grade began, children in one class had no classroom library at all, whereas in another class, the teacher had over 800 books. The books in these classroom libraries were predominantly fiction. This same variability was apparent in subsequent years of the project, as we worked in Grade 3 and then 4. For example, in Grade 4, the average classroom library at the Traditional Instruction school had 276 books, of which 21% were informational books. At the Text Infusion Alone school the average was 169, of which 20% were informational. At the Text Infusion/RFL Instruction school, one teacher had no classroom library and the other teacher had only 24 books, of which 17% were informational.

This limited availability of informational text was similar to data reported in a number of other studies conducted prior to the beginning of the REAL project (Duke, 2000; Moss & Newton, 1998), but recent analyses have yielded similar results. For example, Jeong et al. (2010) inventoried five classroom libraries at each of grades 2, 3 and 4. The percentages of informational text were 22%, 18%, and 19% across the three grade levels. In a study of 318 teachers in K-5 classrooms conducted in 2007-2008 by Ness (2011), teachers reported on average that 33% of the books in their classroom libraries were informational texts. Across grades 2, 3, and 4, means were 36%, 35%, and 37%, respectively, with ranges from 15% to 75%.

Informational books were provided in the REAL text infusion classrooms in two phases. In the fall, we selected books from booklists to appeal to the interests of children at each grade level, and whenever possible to correspond to the science and social studies themes in the county curriculum. Most of these contained expository writing, with text features such as tables of contents, indexes, and captions. After winter break, children and teachers both played a role in selecting additional books. Children received "catalogs" of informational books and placed stickers on their top choices, and, in addition, they indicated what topics they would like to see addressed in other books. Teachers were also given a list of books from which to make selections. As an illustration, we describe here the books added to classroom libraries in fourth grade. Sixty-six informational books were purchased for each classroom, including 61 expository, 4 narrative-informational, and 1 mixed text book. Eleven of the books were selected to reinforce topics in the social studies curriculum. Others featured animals, famous people, sports, nature, and science. Most of the books had text access features. The selection of books included those that were appropriate for students at a range of reading levels.

Student Reading Comprehension in the REAL Project

Reading achievement was measured using a standardized assessment and two researcher-devised assessments. The standardized test was the Gates-MacGinitie Reading Test (GMRT) 4th edition (MacGinitie, MacGinitie, Maria, & Dreyer, 2000). At Level 2, the GMRT has subtests for decoding (word recognition), word knowledge, and comprehension; at Levels 3 and 4 it has vocabulary and comprehension subtests. Although the GMRT includes both fiction and informational passages on its comprehension assessment, scores are not broken down by type of passage. Because of our primary interest in comprehension of informational text, we conducted supplementary analyses where we categorized passages according to text type so that we could compare scores on fiction and nonfiction (Ruetschlin, Finger, & Dreher, 2005). One of the researcher-devised comprehension assessments was based on open-ended written responses to extended informational texts, and the other tapped student competencies in the use of text access features specific to informational text (e.g., indexes and glossaries).

Because of the small scale of the study, with only one school per treatment, and the limited number of teachers within each condition, it was not possible to conduct multi-level analyses. This is a significant limitation because differences among teachers were strong, even within the same school/treatment condition. Moreover, the clustering of children within classrooms varied across the years; that is, students did not remain in intact classes as they moved from grade to grade. The longitudinal analyses are also limited by the decreasing sample size due to attrition from Fall of Grade 2 through Spring of Grade 4.

The GMRT Assessments

In the first year of the study, children took the GMRT in the fall to provide baseline information and again in the spring. In subsequent years, they took it only in the spring. Analyses of variance using extended scale scores (ESS) were conducted, with time of test as the repeated measure and treatment condition (Text Infusion/RFL Instruction, Text Infusion Alone, and Traditional Instruction) as the between-subjects factor. Extended scale scores allow one to compare achievement across the entire range of grade levels tested on the GMRT. A score of 500 represents the normative achievement of a 5th grader at the beginning of the school year; the normal curve equivalent score (NCE) would be equal to 50. The same pattern of results was obtained in the analyses on each of the GMRT subtests and on total reading. We focus here on the Comprehension assessment. Mean scale scores are shown in Table 3. During the first year of the project, in second grade, children improved significantly from fall to spring, as one would expect given regular classroom instruction, $F(1, 176) = 271.72$, $p < .001$, partial $\eta^2 = .61$. Children in all three conditions had similar scores at the outset of the project and showed comparable gains in comprehension. Thus, contrary to predictions, reading achievement as indexed by a widely-used standardized assessment was not greater as a result of enhanced exposure and use of informational books in the classroom during second grade. Results were similar in the longitudinal analyses; children continued to improve in their reading skills through fourth grade, F (103) = 239.45, $p < .001$, partial $\eta^2 = .70$, but with no differential effects related to treatment condition.

In order to better contextualize the reading comprehension abilities of the children in the REAL project, we relate their performance at the end of grade 3 to national norms. This time point corresponds to the "Reading by 9" initiative popularized in the late 1990s that emphasized the need to be ready to transition to "reading to learn" in 4th grade. Children who have a grade equivalent (GE) of 3.9 in comprehension on the GMRT would have an ESS of 477 and an NCE of 52. In the REAL sample, only 25% of the children were reading at or above 3.9 on the spring of Grade 3 assessment. The median GE was only 2.8. Moreover, 25% of the students were reading at or below a GE of 2.3. Clearly, as intended when we selected the district and the schools within it, children's reading comprehension was low relative to national norms.

Table 3. *Extended Scale scores for GMRT Comprehension by Treatment Condition and Grade Level (SDs in Parentheses)*

Treatment	*N*	Fall Gr. 2	Spring Gr. 2	Spring Gr. 3	Spring Gr. 4
Text Infusion/RFL	28	383.71 (34.19)	422.46 (34.27)	445.82 (27.60)	470.50 (21.78)
Text Infusion Alone	48	395.63 (37.99)	416.63 (36.46)	447.15 (38.83)	468.27 (33.23)
Treatment	*N*	Fall Gr. 2	Spring Gr. 2	Spring Gr. 3	Spring Gr. 4
Traditional	30	392.60 (58.08)	422.37 (44.88)	458.07 (30.03)	475.77 (31.69)

To address our specific interest in informational text comprehension, the passages on all four forms of the GMRT comprehension subtests were classified as fiction or nonfiction, with nonfiction further categorized as narrative-informational or expository. (No instances of mixed text were identified.) Two raters independently categorized each passage and established good inter-rater reliability. The classifications were subsequently compared to those reported in the GMRT technical manual. For Levels 2 through 4, the GMRT categorizes passages by content as fiction, social sciences, or natural sciences, and by type as narrative, expository, or setting. They used setting for "passages that seemed characteristic of those sections of stories that do not actually move the account forward in time" (MacGinitie et al., 2002, p. 21), and we classified all the setting passages as fiction.

Our passage classification matched on 36 of the 42 passages (86%) across the four forms. For two of the passages on which there was disagreement (one on each of the Level 2 forms), the GMRT technical manual indicated that the source material was social studies exposition. However, when taken out of context as in an assessment, these passages appeared to be stories and we coded them as such. On Level 2T, two other mismatches occurred for (a) a passage the GMRT considered social studies narrative but appeared to be a story so we coded it fiction, and (b) a passage the GMRT classified as natural science narrative but we classified as expository. On Level 4S, there were two mismatches. In one case, the GMRT classified a passage as social studies narrative, but we classified it as exposition. In the other case, the passage was labeled exposition, but we judged it to be narrative-informational. Although there were few mismatches, those that occurred underscore the difficulty of classifying short segments out of context.

Table 4 shows the number of passages of each type on each form coded according to our system, as well as the number of test items of each type. The proportion of nonfiction passages increases with grade level, as one would expect. Unexpectedly, the two types of nonfiction passages, expository and narrative-informational, were unevenly distributed across forms and grade levels. Proportion correct of each item type was calculated for each student, and these data were used to test for internal consistency reliabilities for each of the item subsets; Cronbach's alphas are presented in Table 4. (In an additional set of analyses not described here, we determined that the nature of the questions changes from predominantly literal in Grade 2 to predominantly inferential in Grades 3 and 4 and that students performed better on literal questions at all four testing points.)

Within-subject analyses of variance were conducted to compare comprehension on the different passage types. Treatment condition was not included as a factor in order to increase power. All children who completed each test were included in the analyses, regardless of how long they had been enrolled at the participating schools. Mean proportions correct on fiction and nonfiction items are shown in Table 5. In the fall of Grade 2, on Form 2S, and in the spring of Grade 2, on Form 2T, students scored significantly better on the fiction questions than the nonfiction, $F\,(1,222) = 223.70$, $p < .001$, partial $\eta^2 = .50$, and $F\,(1, 211) = 42.23$, $p < .001$, partial $\eta^2 = .17$, respectively. Grade 3 Form S data also indicated significantly better performance on fiction than on nonfiction, $F\,(1, 183) = 33.09$, $p < .001$, partial $\eta^2 = .15$. However, in Grade 4 Form T, performance was comparable on the two text types, $F\,(1, 206) < 1$. The commonplace generalization that children comprehend fiction better than nonfiction was supported in Grades 2 and 3, but by Grade 4 informational text was comprehended as well as fiction. This latter outcome contrasts with the 2009 NAEP results, which showed an advantage for literary over informational comprehension for fourth graders. The pattern also varies from that reported by Diakidoy et al. (2005) who found that fourth graders had better scores on expository than on narrative, but second graders were comparable.

Outcomes on reading assessments may vary depending on the nature of the nonfiction passages that are used (e.g., narrative-informational and/or expository). Because the Grade 4 test had sufficient numbers of items of each type, with good internal consistency reliability on each subscale, an additional analysis was conducted comparing comprehension of fiction, narrative-informational, and expository text types. The overall main effect was significant, $F(2,206) = 6.90$, $p < .001$, partial $\eta^2 = .06$, with follow up analyses showing that expository text comprehension was significantly better than comprehension of narrative-informational text; comprehension of fiction was at an intermediate level, not significantly different from either of the other two text types. A parallel analysis was conducted for Grade 2 (Form S), but results should be interpreted cautiously given the low reliability of the expository text subscale. This analysis also revealed an overall effect of text type, $F\,(2,222) = 129.64$, $p < .001$, partial $\eta^2 = .37$. Comprehension was best on fiction, with narrative-informational text in the middle, and with expository text comprehension the weakest; all pairwise comparisons were significantly different. These data are also shown in Table 5.

Several issues must be kept in mind in interpreting these results, the most important of which is that the passages were very short, averaging from 82 words in the Grade 2 tests to 93 words in Grade 4. Classification as to text type is more accurate with longer passages. Moreover, students reading short expository passages often do not encounter the same text features they would encounter in longer expository texts, such as bold-faced words, headings, and glossaries. Also, complex organizational text structures are not as evident in short expository passages as they are in longer expository texts. A final limitation of this analysis is that the developers of the GMRT did not design the test to yield separate scores on comprehension of fiction and non-fiction. Although we found acceptable internal consistency reliability on all but one of the subscales, we were not using the test as the developers intended.

Table 4. *Number of Passages and Items of Different Text Types on the GMRT Comprehension Test and Internal Consistency Reliabilities of the Derived Sub-scales*

Form	Passage Type			
	Fiction	Overall Nonfiction	Narrative Info	Expository
Grade 2 (S)	6	4	2	2
Grade 2 (T)	6	4	0	4
Grade 3 (S)	6	5	0	5
Grade 4 (T)	4	7	3	4
	Item Type			
	Fiction	Overall Nonfiction	Narrative Info	Expository
Grade 2 (S)	24	15	8	7
Grade 2 (T)	24	15	0	15
Grade 3 (S)	29	19	0	19
Grade 4 (T)	15	33	17	16
	Cronbach's Alpha			
	Fiction	Overall Nonfiction	Narrative Info	Expository
Grade 2 (S)	.87	.75	.78	.46
Grade 2 (T)	.87	.79	..	.79
Grade 3 (S)	.86	.80	..	.80
Grade 4 (T)	.79	.85	.70	.70

Table 5. *Proportion of Correct Responses (SDs in parentheses) on the Fiction and Nonfiction Items of the GMRT Comprehension Test*

		Type of Test Item			
Form	*N*	Fiction	Overall Nonfiction	Narrative Info	Expository
Grade 2 (S)	223	.53 (.25)	.36 (.22)	.41 (.27)	.30 (.22)
Grade 2 (T)	212	.61 (.23)	.54 (.25)	..	.54 (.25)
Grade 3 (T)	184	.51 (.24)	.45 (.23)	..	.45 (.23)
Grade 4 (S)	207	.51 (.24)	.51 (.20)	.48 (.22)	.52 (.21)

The analysis of passage types on the GMRT points to an important issue with regard to nonfiction text: how it is categorized. In particular, the distinction between expository and narrative-informational text is often problematic. For example, although the intent of the GMRT developers was that Forms 2S and 2T present children with equivalent numbers of expository and narrative-informational passages, we found disparity between the forms. Our classification indicates that each form has 15 nonfiction items; but 2T has 15 items on expository passages, whereas 2S has 7 items on expository passages and 8 on narrative-informational. As noted, part of the difficulty may be that these passages are not only very short, but that they appear out of context. But to complicate matters, nonfiction is not categorized consistently in the literature. In the classification system used on the 2009 NAEP, narrative-informational passages are included within the literary category, and expository passages are included within informational (NCES, 2009). In contrast, the Common Core State Standards include both types in their definition of informational text (CCSS, 2010). This variability in classification schemes limits generalizability across studies.

It is clear from our analyses that the distinction between narrative-informational and expository text is important, and these text types should be examined separately. Moreover, it is not yet understood whether children process narrative-informational texts more similarly to fictional narratives or to nonfictional exposition. However, research reveals differences in the competencies that contribute to comprehension of the different genres. Best, Floyd, and McNamara (2008) found that among third graders, comprehension of narrative text was most influenced by decoding skills, whereas comprehension of expository text was most influenced by world knowledge. The study provides confirmation of the widely-stated assertion that children need a solid knowledge base in order to understand expository text.

Reading-and-Writing-to-Learn Assessments

To more directly tap the comprehension skills that were the focus of the REAL project, performance assessments were developed for each grade level and were pilot tested prior to implementation. For second and third grade, two alternate form expository texts were created and were presented in counterbalanced order across fall and spring test sessions. The second grade texts dealt with either snakes or frogs, and each was bound in an 8-page booklet with color illustrations and a table of contents. The third grade texts dealt with either Rome or Egypt. Booklet length was increased to 16 pages; the texts included color illustrations and text access features (a table of contents, glossary, and index). For fourth grade, the task increased in complexity with children receiving two booklets at each testing session, requiring them to integrate information across booklets. Fourth graders received booklets on either Boston and Chicago, or Hawaii and Alaska, presented in counterbalanced order across fall and spring testing. The four booklets for fourth grade were each 18 pages long, including illustrations, a table of contents, glossary, and index. At all three grade levels, children responded in writing to open-ended prompts about the materials they read, as described below.

The Reading and Writing to Learn assessments differed across the years because of children's increasing competencies. Although we developed this assessment prior to the publication of the Common Core State Standards, the expository text comprehension demands at each grade level were consistent with the standards. Table 6 shows the specific prompts that were used in each assessment and examples of the qualitative scoring rubrics that were used.

Grade 2 Performance Assessment. Students were presented with one of the two books and an accompanying response sheet. Children's responses to the prompts shown in Table 6 were scored according to a 5-point scale ranging from 0 to 4. Portions of the rubric for question 2 are also shown. Each response was scored by two independent coders; reliability was strong, with correlations averaging about .90. The two questions were summed for a total score. Although the scores were significantly correlated, internal consistency reliability was rather low, .52 in the fall and .47 in the spring.

Table 6. *Prompts and Sample Scoring Rubrics for the Reading and Writing to Learn Performance Assessments*

Grade 2

Prompt

1. Use your book to learn new information about frogs. Find out interesting information about frogs. Write down what you have learned on the lines below.

Today I learned about frogs. Here are some interesting things that I learned. …

2. Pretend your friend wants to pick one of these frogs as a pet. Use your book to find out about these two frogs. Which frog should your friend pick as a pet? Explain why your friend should pick that frog.

Circle the frog your friend should pick (photographs were provided of the two frogs described in the text).

My friend should pick this frog as a pet because …. .

Portions of the Scoring Rubric for Question 2

5 points: Answer is written in student's own words, claims one of the animals as the best choice, uses information from more than one place in the text to support the choice, and includes a contrasting of the two animals to further explain and justify the choice.

1 point: answer is expressed as an opinion, or uses only picture clues, but does not include information from the text.

Grade 3

Prompt

Pretend that you are going back in time to visit the people of Rome/Greece long ago. Write what the people would be like. Be sure to include information on how they would dress, what they would eat, what they do for fun, and any other information that would help you tell us what their life would be like.

Portions of the Scoring Rubric for the Use of Notes

4 points: Notes are clear and serve as the primary basis for the details in the writing and the organizational structure.

1 point: Notes are recorded but there is little or no use of notes in the writing.

Table 6 (Continue). *Prompts and Sample Scoring Rubrics for the Reading and Writing to Learn Performance Assessments*

Grade 4

Prompt

Where do we go?

Today you will pretend your class has won an all-expense paid trip to 1 of 2 exciting places. Your job is to research both locations and decide which place you think your class should visit. You will have two short books to read about these places.

We will give you time to read the first book and take notes about what you like and don't like. Then we will give you time to read the second book and take notes. You will use the information you have gathered to choose where you think your class should go.

After you have decided where you think your class should go, you will write a letter to your teacher persuading him or her to take your class to this place. Be sure to include information from the 2 books to support your choice.

Portions of the Scoring Rubric for Persuasion

4 points: The student takes a clear stand on where the class should go and fully supports it with reasons based on the reading and their prior knowledge. The student details why they should go to the chosen place, and why they should not go to the other place.

1 point: The student does not take a stand on the issue. The student presents some information but it is not clear where he/she would prefer to go, or why they came to his/her decision.

Table 7. *Mean Scores on the Reading and Writing to Learn Assessments (SDs in Parentheses) by Treatment Condition, Grade Level, and Assessment Time*

Grade	Treatment Condition		
	Text Infusion/RFL Instruction	Text Infusion Alone	Traditional Instruction
2	*n = 47*	*n=67*	*n=65*
Fall	3.13 (1.41)	3.30 (1.76)	3.31 (1.89)
Spring	4.13 (1.75)	4.18 (1.57)	4.14 (1.94)
3	*n=33*	*n=52*	*n=35*
Fall	12.45 (3.48)	12.58 (4.23)	13.69 (5.02)
Spring	14.64 (3.34)	14.73 (4.24)	16.60 (5.01)
4	*n=27*	*n=30*	*n=31*
Fall	11.19 (2.94)	11.23 (2.81)	11.35 (3.05)
Spring	10.56 (3.03)	11.87 (3.14)	11.94 (3.35)

Table 7 shows the mean scores on the Reading and Writing to Learn assessment in the fall and in the spring for each treatment condition. Preliminary analyses revealed that scores on the two alternate forms were comparable and that scores were comparable across conditions at baseline. A repeated measures analysis of variance revealed significant growth over the school year F (1, 176) = 41.24, p < .001, partial η^2 = .19 (fall mean = 3.26 and spring mean = 4.15). Contrary to predictions, however, children who received the text infusion with reading for learning instruction did not outperform children experiencing only the text infusion or traditional classroom instruction, even though this assessment was more closely aligned with the focus of the intervention than was the GMRT.

Grade 3 Performance Assessment. In Grade 3, children responded to a single elaborated question, shown in Table 6. Along with the passages, they were provided with a grid for taking notes to help them prepare their response. The worksheet included spaces for recording information about clothes, food, dress, and other information. Responses were scored according to four criteria: topic use, use of notes, accuracy of information, and quality of writing. Scores on each could range from 0 to 4. See Table 6 for a portion of the rubric for use of notes. Each response was scored by two independent coders, and inter-rater reliability was strong, with correlations above .90. The four separate scores in the coding rubric were summed for a total score. Cronbach's alpha revealed good internal consistency reliability (.78 in the fall and .75 in the spring).

Mean scores are shown in the middle section of Table 7. Analysis of variance revealed significant improvement from fall to spring, F (1, 117) = 24.24, $p < .001$, partial $\eta^2 = .17$, with a mean score of 12.87 in the fall and 15.25 in the spring. However, again contrary to predictions, children in the Text Infusion/RFL Instruction condition did not achieve higher comprehension scores than children in the other two conditions, nor was there differential growth over the school year.

Grade 4 Performance Assessment. The Grade 4 task required students to integrate information across two expository texts to derive their responses to a single question (see Table 6). Along with the passages, students were given a response sheet with columns for taking notes about what they liked and did not like about each of the two possible destinations for a class trip. A 5-element coding rubric was used for the responses, with 0 to 4 points on each element. Students were scored on persuasiveness, organization, style, grammar, and use of notes. A portion of the rubric for persuasion is shown in Table 6. Each response was scored by two independent coders, with reliability again strong (correlations among raters averaged .88). The five elements in the coding rubric were summed for a total score. Cronbach's alpha revealed good internal consistency reliability (.82 in the fall and .85 in the spring).

On this more challenging performance assessment, students showed no significant improvement from fall to spring, F (1, 85) < 1. The mean score in the fall was 11.26 and in the spring it was 11.45. Once again, contrary to predictions, informational text infusion in the classroom, with or without reading for learning instruction, did not affect performance (see Table 7).

Predictors of Performance on the Reading and Writing to Learn Assessments. As was just demonstrated, a key question that motivated the REAL project was answered negatively; that is, the Text infusion/RFL Instruction intervention did not promote growth over time in children's comprehension of expository text on the RWTL assessment. Another important question concerns the role of early literacy competencies on subsequent achievement. As has been amply shown in the literature, children come to school with vast differences in background knowledge that are not easily supplemented in the classroom, and the comprehension skills that children have acquired by the end of first grade are strong predictors of performance throughout their school years (Duke & Carlisle, 2011; Neuman, 2006). Multiple regression analyses were therefore conducted to examine the extent to which competencies at the beginning of Grade 2 predicted growth on the performance assessments from fall to spring in each of Grades 2, 3, and 4. If children's early facility in reading and writing about informational text continues to predict performance in subsequent years, above and beyond comprehension as measured by the GMRT, it would provide evidence that further efforts to enhance children's experiences with informational text are indeed warranted.

In each analysis, students' scale scores on the Fall of Grade 2 administration of the GMRT were entered into a regression analysis, treating Decoding, Word Knowledge, and Comprehension as separate variables, along with the Fall of Grade 2 scores on the Reading and Writing to Learn Assessment. For the regression predicting Spring of Grade 2 RWTL scores, no other predictor variables were entered. For the regressions of Spring of Grade 3 and 4 RWTL scores, the Fall RWTL scores were used for the respective grade level in order to test for growth across the year. The results of the three analyses are shown in Table 8.

The overall model for Grade 2 was significant, F (4, 172) = 15.97, $p < .001$, accounting for 27% of the variance. Significant variance was accounted for by the Fall RWTL scores, by GMRT Decoding, and GMRT Word Knowledge. The GMRT Comprehension scores did not contribute additional variance beyond that accounted for by the other variables. At this early grade level, students' entering decoding skills best predicted growth in the performance assessment. The overall model for Grade 3 was also significant, F (5, 112) = 4.59, $p < .001$, accounting for 17% of the variance. However, no individual predictors were significant, with GMRT Comprehension the only one that even approached significance at $p < .10$.

Of most interest was the longer-term analysis of predictors of reading and writing to learn at the end of Grade 4. This analysis provided clear evidence of the importance of early informational text competencies. The model was significant, F (5, 78) = 18.44, $p < .001$, accounting for 54% of the variance. The Fall of Grade 2 RWTL score was the strongest individual predictor of RWTL growth in Grade 4. The beta weight was even stronger than that of the Fall of Grade 4 RWTL score. None of the early GMRT scale scores significantly predicted unique variance, although Decoding came close at $p<.07$. Thus, the children who at the beginning of Grade 2 already had the comprehension skills needed to read an expository text and respond in writing to open-ended prompts fared better on the end-of-grade 4 RWTL assessment. We believe that Question 2, which called for the child to justify the choice of a pet for a friend, was a particularly sensitive assessment of comprehension because one of the animals was described as harmless and the other as poisonous.

Table 8. *Regression Analyses of Early GMRT Scores and Reading and Writing to Learn (RWTL) Scores as Predictors of Growth in RWTL in Grades 2, 3, and 4*

	B	*SE B*	Beta	*t*	Sig.
Spring Grade 2 RTWL					
Intercept	-.18	1.41		-.127	.90
Fall Gr. 2 RWTL	.33	.07	.33	4.54	<.001
Fall Gr. 2 Comprehension	.003	.004	.07	.81	.42
Fall Gr. 2 Decoding	.02	.01	.44	3.42	.001
Fall Gr. 2 Word Knowledge	-.01	.01	-.25	-2.03	.04
Spring Grade 3 RTWL					
Intercept	-1.40	5.28		-.27	.79
Fall Gr. 2 RWTL	.31	.25	.12	1.23	.22
Fall Gr. 2 Comprehension	.02	.01	.19	1.70	.09
Fall Gr. 2 Decoding	.02	.02	.16	.88	.38

Note. R^2 = .27 for Grade 2, R^2 = .17 for Grade 3, and R^2 = .54 for Grade 4.

Table 8 (Continue). *Regression Analyses of Early GMRT Scores and Reading and Writing to Learn (RWTL) Scores as Predictors of Growth in RWTL in Grades 2, 3, and 4*

	B	*SE B*	Beta	*t*	Sig.
Spring Grade 2 RTWL					
Intercept	-.18	1.41		-.127	.90
Fall Gr. 2 RWTL	.33	.07	.33	4.54	<.001
Fall Gr. 2 Comprehension	.003	.004	.07	.81	.42
Fall Gr. 2 Decoding	.02	.01	.44	3.42	.001
Fall Gr. 2 Word Knowledge	-.01	.01	-.25	-2.03	.04
Spring Grade 3 RTWL					
Intercept	-1.40	5.28		-.27	.79
Fall Gr. 2 RWTL	.31	.25	.12	1.23	.22
Fall Gr. 2 Comprehension	.02	.01	.19	1.70	.09
Fall Gr. 2 Decoding	.02	.02	.16	.88	.38
Fall Gr. 2 Word Knowledge	.001	.02	.004	.03	.98
Fall Gr. 3 RWTL	.07	.10	.07	.67	.50
Spring Grade 4 RTWL					
Intercept	1.61	3.14		.51	.61
Fall Gr. 2 RWTL	.92	.19	.44	4.86	<.001
Fall Gr. 2 Comprehension	.01	.01	.09	.82	.42
Fall Gr. 2 Decoding	.03	.01	.35	1.86	.07
Fall Gr. 2 Word Knowledge	-.02	.02	-.25	-1.50	.14
Fall Gr. 4 RWTL	.27	.10	.24	2.64	.01

Use of Informational Text Features

Children's ability to access information using text features was assessed in the first two years of the project. Text access was a central component of the instruction teachers in the Text Infusion/Reading for Learning condition were asked to provide. Considerable attention was given to text access features in the in-service training, and texts were selected for the classroom libraries based on the quality of these features.

Samples of the text access items are provided in Table 9. The same assessment instrument was used in the fall and spring. In the Grade 2 task, children were asked six questions about finding information in a table of contents and an index. In Grade 3, a more demanding assessment was used, consisting of 12 questions, most of which required children to decide whether the requested information could be derived from the table of contents, the index, or the glossary. Text access was not examined in fourth grade because many students scored at or near ceiling on the third grade measure. The decision to focus only on the primary grades turned out to be consistent with the subsequently-disseminated Common Core standards, which specifically address use of text access features in Grades 2 and 3, but not in Grade 4.

Table 9. *Sample Items from the Text Access Tasks*

Grade 2	
TABLE OF CONTENTS	
What Is a Bear?	5
Where Do Bears Live?	8
What Do Bears Eat?	13
Baby Bears	20
Polar Bears	25
Brown Bears	29
Black Bears	35
Bears In Danger	38
Bears Are Wild Animals	43
Words You Should Know	46
Index	47
1. Which page would tell you what bears eat? _______	
1. Which page would tell you about brown bears? ______	
Grade 3	
Here are the table of contents, glossary, and index from a book about France.	
(Materials are provided to the students.)	
1. On what page does the chapter "Vacation Time" begin? _____	
2. On what page can you find information on castles? _____	
3. What does "immigrant" mean? _____	

Table 10 provides the mean proportion of items answered correctly on the text access tasks in Grades 2 and 3. The Grade 2 repeated measures analysis of variance revealed significant improvement from fall to spring, $F(1, 173) = 111.99$, $p = .001$, partial $\eta^2 = .39$. Of most importance to the goals of the study, and consistent with predictions, was the presence of a reliable interaction of time of test and treatment condition, $F(2, 173) = 3.09$, $p = .048$, partial $\eta^2 = .03$. Children in the Text Infusion/RFL Instruction condition made relatively greater gains over the school year than children in either of the other conditions. In Grade 3, repeated measures analysis of variance revealed gains from fall to spring, $F(1, 158) = 17.90$, $p < .001$, partial $\eta^2 = .10$, but no differential improvement across conditions.

Across all three project years, scores on the GMRT were significantly correlated with text access scores. Similarly, scores on the Reading and Writing to Learn tasks were significantly correlated with use of informational text features.

Motivation and Voluntary Reading Activity

Because the REAL project was conceptualized within the engagement model of reading (Baker, Dreher, & Guthrie, 2000), we examined not only reading achievement, but also enablers of that achievement. In this view, students who are motivated to read and who choose to read frequently will be better comprehenders than those who are less motivated and who read less. It was hypothesized at project outset that increasing student access to interesting texts would increase reading motivation and reading activity, which in turn would increase reading comprehension. We briefly report the results of our analyses of motivation and reading activity, but without the level of detail given to comprehension.

Table 10. *Mean Proportion Correct on Text Access Tasks in Grades 2 and 3 by Treatment Condition and Time of Test*

Grade	Text Infusion/RFL Instruction	Text Infusion Alone	Traditional Instruction
2			
Fall	.44	.51	.57
Spring	.82	.84	.78
3			
Fall	.66	.77	.78
Spring	.75	.80	.81

The Motivations for Reading Questionnaire – Primary (MRQ-P), a shorter and simplified form of an instrument designed for older elementary school children (Baker & Wigfield, 1999; Wigfield & Guthrie, 1997) was administered in the fall and spring of Year 1 and in the spring only in Years 2 and 3. Students rated their motivation with respect to different aspects of reading on a 3-point scale. Item ratings were summed for a total motivation score, and subscale scores for four dimensions were also obtained: perceived competence, interest (intrinsic motivation), recognition (extrinsic motivation), and social interaction. Analyses of internal consistency reliability were conducted each time the MRQ-P was administered, and alphas were at acceptable levels (typically around .85). Sample items are shown in Table 11.

In Grade 2, children showed a significant *decline* in overall motivation from fall to spring, consistent with a number of other studies showing declines in the early grades. However, longitudinal analyses did not reveal further declines in third and fourth grade. Similar to the results for the reading assessments, motivation was not impacted differentially by classroom text infusion. Analyses of the separate motivation subscales revealed generally similar patterns to the overall scale, with the exception of the perceived competence subscale. Although children's perceived competence in reading declined from fall to spring of second grade, it increased over the next two years.

Table 11. *Sample Items from the Motivation for Reading Questionnaire (MRQ-P)*

1. Perceived Competence			
How good a reader are you?			
	Very good	OK	Not very good
2. Interest (intrinsic motivation)			
How much do you like to read about new things?			
	Very much	A little	Not much at all
3. Recognition (extrinsic motivation)			
How much do you like to get praise for your reading?			
	Very much	A little	Not much at all
4. Social			
How much do you like to tell your family about what you are reading?			
	Very much	A little	Not much at all

Note. Responses at the positive end of the scale were scored as 3, those at the negative as 1.

Correlational analyses examined whether higher levels of reading motivation were associated with better reading comprehension. Motivation scores were associated with performance on the GMRT over the years, with the strength of the correlations greater for particular subscales. For example, in Grade 4, perceived competence and reading for recognition (extrinsic motivation) were significantly correlated with achievement, but reading for interest (intrinsic motivation) and reading for social interaction were not. Motivation scores were more weakly associated with performance on the Reading and Writing to Learn assessment than the GMRT, especially for the assessments administered in the first year of the project. Children's motivation was also weakly associated with performance on the text access tasks, with stronger relations for the perceived competence subscale than for the overall measure or other subscales.

Table 12 shows the consistent relations between perceived competence and GMRT reading comprehension over the years. It is particularly noteworthy that as early as the beginning of second grade, children's perceptions of themselves as readers were associated with their actual performance, concurrently and into the future. Also noteworthy is the suggestion of bidirectional influences; Fall of Grade 2 perceived competence was associated with Grade 4 comprehension, and Fall of Grade 2 comprehension was associated with Grade 4 perceived competence.

Table 12. *Correlations between Perceived Competence and GMRT Comprehension*

Perceived Competence	Reading Comprehension			
	Fall Gr. 2	Spring Gr. 2	Spring Gr. 3	Spring Gr. 4
Fall Gr. 2	.24	.21	.23	.25
Spring Gr. 2	.27	.31	.28	.28
Spring Gr. 3	.25	.27	.22	.21
Spring Gr. 4	.38	.33	.36	.37

Note. Listwise N=101. All correlations are significant at $p < .05$ or better.

The Reading Activity Inventory – Primary (RAI-P), a simplified version of the instrument designed for older children (Wigfield & Guthrie, 1997) was administered following the MRQ-P. Students responded to two types of questions regarding their reading activity out of school, one asking how often they read particular types of books and the second asking whether they had read that type of book within the past week. The RAI-P assesses the amount and breadth of students' reading by examining the frequency with which students read texts of different genres (e.g., mysteries, biographies, nature books). Table 13 shows sample questions for fiction and for informational text.

In Grade 2, analyses revealed significant declines in self-reported reading activity from fall to spring; these declines occurred for children in all treatment conditions. The longitudinal analysis showed that reading activity scores did not decline further in Grades 3 or 4. However, the effect of treatment condition was statistically significant, $F(2, 109) = 5.85$, $p = .004$, $\eta^2 = .10$. Consistent with one of the REAL hypotheses, students who received only traditional instruction reported less frequent outside reading over the years than students who experienced text infusions in their classrooms. (For Traditional Instruction, M = 2.09, SD = .54; for Text Infusion Alone, M = 2.46, SD = .56, and for Text Infusion/RFL Instruction, M = 2.30, SD = .58). Note that a mean rating of 2.0 corresponds to reported reading of a particular type of book "about once a month" whereas a mean rating of 3.0 corresponds to "about once a week." These results suggest that exposing children to interesting informational text in the classroom may stimulate more frequent out-of-school reading.

Table 13. *Sample Items from the Reading Activity Inventory (RAI-P)*

Question Regarding Fiction			
Did you read a make-believe story or book last week (like a mystery or an adventure for your own interest or for fun?			
No	Yes		
How often do you read make-believe stories or books for your own interest or fun?			
Almost never	About once a month	About once a week	Almost every day
Question Regarding Informational Books			
Did you read a nature book last week for your own interest or for fun?			
No	Yes		
How often do you read nature books for your own interest or for fun?			
Almost never	About once a month	About once a week	Almost every day

Note. Responses of "almost every day" were scored as 4; those of "almost never" as 1.

Self-reported reading activity was consistently related to reading motivation, as is the case when the full versions of the MRQ and the RAI are used (Baker & Wigfield, 1999; Beall, 2011; Wigfield & Guthrie, 1997). However, self-reported reading activity was not associated with reading comprehension in this study, across any of the years. This contrasts with other research showing relations between out-of-school reading and reading achievement (Beall, 2011; Serpell, Baker, & Sonnenschein, 2005, Wigfield & Guthrie, 1997).

Contextual and Individual Differences

Opportunities to Read Informational Text in the Classroom

A central premise of the REAL project was that reading informational books would increase the likelihood that students would acquire the knowledge and motivation needed to become better comprehenders. An examination of what took place in REAL classrooms was therefore important as an implementation check and as a means of understanding how contextual factors might impact children's comprehension of informational text (Katenkamp, Garrett, & Baker, 2005).

Classroom observations were conducted in all of the REAL classrooms from October to May during each year of the project, usually three weeks apart. Because the use of informational text could occur at any time during the day, rather than being limited to reading instruction, these observations took place for the full school day. Observers used an adapted version of the Center for the Improvement of Early Reading Achievement (CIERA) School Change Classroom Observation Scheme (Taylor & Pearson, 2001). This scheme includes seven major categories that reflect who is leading the class, the kind of classroom grouping, the major focus of the lesson, the specific type of activity going on, the type of material being used, the nature of the teacher interaction, and the expected pupil response. In addition, the major focus of instruction (i.e., reading, math, science, social studies), the manner in which students were grouped (i.e., whole group, small group, individual), the number of students engaged in the activity, and the total number of the students in the room were recorded. A five-minute coding interval was followed by a five-minute break during which observers continued to take field notes.

We focus here on selected components of the observations. An overall reading composite was created that consisted of how often the coded activity reflected reading or listening to connected text. Word or sentence level activities were not included. The amount of classroom time spent in reading increased across grades two through four. More reading took place in text infusion classrooms; effects were significant in Text Infusion/RFL

Instruction classrooms in Grades 2 and 4 and in Text Infusion Alone classrooms in Grade 3. However, the amount of time spent reading in the classroom was not related to reading achievement or motivation. Increased access to informational books may thus have contributed to increased reading, but increased reading alone was not sufficient to impact comprehension.

Opportunity to listen to the reading of connected text was also examined. Given that teachers in the Text Infusion/RFL Instruction condition were explicitly asked to read aloud the informational books that were provided, it was expected that these teachers would spend proportionally more time in such activity. This expectation was met in second grade, but not in third or fourth. In second grade, teacher read-alouds comprised 20% of classroom reading activity in the Text Infusion/RFL Instruction condition, but only 9% in the Text Infusion Alone and 8% in the Traditional Instruction condition. Proportions decreased to less than 10% in all conditions by 4th grade. The amount of time teachers read aloud to their students was unrelated to reading achievement or motivation.

Opportunities to read different genres (stories and informational texts) were examined. A composite measure of all informational texts was created; this composite included both text and trade books classified as expository, narrative- informational and mixed texts. The percentage of materials used that were coded as informational texts increased significantly from 2nd to 4th grade, from 11% to 20% to 26%. Students in the text infusion classrooms used informational texts more often than those who were in the traditional instruction classrooms, as would be expected given the nature of the intervention. The difference was most dramatic in second grade, where informational texts were used 23% of the time in the Text Infusion/RFL Instruction classrooms but only 5% and 6% of the time in the other two types of classrooms. Contrary to predictions, greater use of informational text was not related to reading comprehension.

Individual Preferences for Informational vs. Narrative Text

Research has shown that young children respond positively to non-narrative text. For example, when Mohr (2003) gave first graders the opportunity to select a book to keep, 84% selected a nonfiction book. Relevant information was collected in the REAL project through reading logs and preference inventories (Beall, Morse, Baker, & Dreher, 2005). We focus here on fourth-grade findings, when students were asked to list their two favorite books for the year from any they had read at school or at home, to explain why they liked these books, and to select a book to keep as a thank-you gift for project participation.

Overall, and in contrast to the results of Mohr (2003), the majority of students listed a narrative book as their first and second favorite. Informational books were the first choice for 26% of the students and the second choice for 33%. Of particular note is that treatment condition was associated with the type of book chosen as a favorite. A greater percentage of students in the Text Infusion/RFL Instruction and Text Infusion Alone conditions listed an informational book as their first favorite than in the Traditional Instruction condition (41% and 31%, respectively, versus 10%), and the same pattern held for the second favorite (44% and 45% versus 17%), suggesting that increasing informational text availability can influence preferences.

The reason most frequently given by students for selecting a particular book as a favorite was emotional appeal (33%; e.g., “It was funny”). Personal interest was given as a reason by 19% of the students (e.g., “It has my favorite animals,” and “It had something to do with my hobby”). Value to the reader was mentioned by 18% of the students (e.g., “It told me many things about cats,” “It shows the basic steps how to play).” Students whose favorite books were informational more frequently gave reasons related to the value of the topic. Guthrie et al. (2007) also found that fourth graders' motivation for reading expository text focused primarily on acquiring new knowledge.

Whether students identified informational books as their favorite was related to reading achievement in the Text Infusion/RFL Instruction condition. Students who selected an informational book as their favorite scored significantly higher on the Word Knowledge/Vocabulary subtest of the GMRT than students who selected a narrative book as favorite. Although student preference for a particular genre was not associated with reading comprehension per se, the fact that it was associated with vocabulary is noteworthy, given the importance of academic vocabulary to reading comprehension.

That the Text Infusion/RFL Instruction intervention had an impact was also suggested by the fact that students who chose an informational book as their favorite reported significantly higher perceived competence, interest, and overall motivation. In addition, students in this treatment condition who identified an informational book as a favorite reported more frequent out-of-school reading and more diversity of reading genres. No relations among these engagement variables were evident in the other two conditions.

Gender Differences in Reading, Engaging, and Learning

In this section we address gender differences on the various outcome measures in the REAL project. Gender is an individual difference variable of particular importance to this topic because it has often been argued that boys find informational text more appealing than do girls. If this is the case, one might expect boys to have differentially better comprehension of informational text than narrative text, or to be more motivated to read informational books than fiction. Oakhill and Petrides (2007) tested the possibility that the reason why boys in England showed a large reading comprehension increase from one administration of a national reading test to another is because the passage topic was more appealing to them. The topic the previous year had been about leaving home in wartime and that year it was about spiders. Using a within-subjects design, they found that boys overwhelmingly preferred the passage about spiders and girls the passage about wartime evacuation. Of particular relevance is that boys had better comprehension of the passage about the preferred topic, whereas girls had equally good comprehension of both passages.

Reading Achievement. On major large-scale reading assessments, both national and international in scope, girls consistently outperform boys. For example, on the 2009 NAEP, fourth-grade girls had significantly higher scale scores than boys, and fewer scored below the basic level (30% vs. 36%). On the 2006 PIRLS, fourth-grade girls had significantly higher scores than boys in all but two of the 46 participating educational systems around the world (Mullis et al., 2007).

Within the REAL project, reading achievement on the GMRT was consistently better for girls than for boys, across years and across GMRT subscales. The advantage was not statistically significant in grade two, but it was in grades three and four. Across the full longitudinal sample, the main effect of gender on GMRT comprehension was significant, F (1, 104) = 10.55, p = .002, partial η^2 = .09. The mean extended scale score for girls at the end of fourth grade was 458.58 whereas for boys it was 441.43.

We also tested the possibility that the gender difference in reading comprehension might be attenuated when informational texts are used as opposed to fictional. Using our constructed GMRT subscales, we found no evidence of an interaction of gender with text type, either when we compared informational vs. fiction, or when we compared expository and narrative-informational text types separately. In contrast, on the NAEP, girls had significantly better scores on both literary and informational text than did boys (see Table 1).

The advantage for girls on literary text was 8 scale points, but on informational text it was only 6.

On the Reading and Writing to Learn performance assessments, girls and boys performed comparably in second grade. Girls scored higher than boys during third grade but the effect of gender was not statistically significant ($p = .07$), means = 14.60 and 13.75, respectively. The gender difference was statistically significant in Grade 4, $F(1, 84) = 5.69$, $p = .012$, partial $\eta^2 = .06$, with mean scores of 12.10 for girls and 10.78 for boys.

Although considerable evidence shows girls outperforming boys on tests of reading, we are not aware of any studies that have specifically addressed gender differences in information access. Performance on the REAL text access tasks was better for girls than for boys, with statistically significant advantages in both years (second grade, $F(1,173) = 5.82$, $p = .02$, partial $\eta^2 = .033$ and third grade, $F(1, 155) = 7.84$, $p = .006$, partial $\eta^2 = .048$). Mean proportions correct in Grade 2 were .71 for girls and .61 for boys, and In Grade 3, .78 and .69, respectively.

Reading motivation, reading activity, and book preferences. Gender differences in motivation are often reported in the literature. For example, Baker and Wigfield (1999) found consistently higher scores for fifth- and sixth-grade girls on the Motivations for Reading Questionnaire. However, in the REAL project, boys and girls generally had comparable levels of motivation across all three years. Similarly, there were no differences in perceived competence, despite the fact that girls earned objectively higher reading achievement scores than did the boys. At the final assessment point, at the end of Grade 4, girls had a mean score of 2.57 and boys 2.47 out of 3 possible. Clearly, all students had a positive appraisal of their competencies. Girls typically report more frequent out-of school reading than boys (Beall, 2011; Baker & Wigfield, 1999), but gender differences were absent on the reading activity inventory in the REAL project.

Gender differences are often reported in the types of books students prefer to read. For example, Mohr (2003) found that 96% of first-grade boys selected nonfiction books for themselves, whereas 69% of the girls did so. In addition, Oakhill and Petrides (2007) reported that an unpublished survey of 10-11-year-olds found that boys preferred factual books and girls preferred storybooks. In contrast, Chapman, Filipenko, McTavish, and Shapiro (2007) found that first graders showed no gender differences in their preferences for stories or informational books when choosing books for themselves. But when asked about what other children would prefer, both boys and girls thought that boys would prefer informational books whereas girls would prefer stories. In the REAL project as well, no gender differences were found in the types of books that were identified as favorites, indicating that informational books can be equally appealing to both boys and girls. However, when given a choice of which book they would like to receive as a "thank you" gift at the end of the project, more boys than girls chose an informational book, but overall, stories were chosen by 68% of the children.

Summary and Conclusions

The REAL project was designed to investigate whether a classroom intervention that enhances young children's experience with informational books would increase reading achievement and engagement. Contrary to expectations, the intervention had minimal impact on student outcomes. Students in second grade improved in GMRT reading comprehension from fall to spring to the same extent regardless of whether they had additional informational books in their classrooms and specific instruction in their use, and the same pattern obtained in grades three and four. The RWTL performance assessment, administered in both fall and spring every year, revealed improvement over the year in second and third grades, but no differential progress related to the intervention. The one significant treatment effect was obtained on the text access task, which most closely parallels the instruction teachers were asked to provide. Students in the Text Infusion/RFL Instruction condition made greater gains in their ability to use text access features than students in the other two conditions. However, the treatment effect was found only in second grade, not in third.

Students declined in reading motivation during the course of second grade and into third grade, again regardless of intervention condition. Motivation stabilized, showing no further declines in fourth grade. Self-reported reading activity also declined over the first year of the project and stabilized subsequently. Of importance is that students with increased access to informational books in the classroom reported more out-of-school reading across the years than students without the increased access.

By the end of the project in fourth grade, the infusion of informational books into the classroom libraries appears to have influenced student preferences for reading material. Students in both text infusion treatments were more likely to select an informational book as one of their two favorites for the year. Students who selected informational books tended to do so because they satisfied a desire for information or taught the students something that was important to them. Students who expressed a preference for informational books at the Text Infusion/RFL Instruction school may also have experienced an increased interest in reading, as evidenced by more self-reported reading activity out of school.

Despite the lack of evidence of improved comprehension in the REAL project that can be attributable to the text infusion intervention, it should not be concluded that students will not benefit from increased access and instruction in using informational books in the early grades. Substantial correlational evidence exists that should not be ignored, both within our own study and external to it. Moreover, since the REAL project was conceptualized a decade ago, several empirical studies now provide evidence that interventions designed to increase children's understanding and use of informational text result in growth in reading comprehension (e.g., Guthrie, Wigfield, Barbosa et al., 2004; Guthrie et al., 2007; O'Hara, 2007; Williams et al., 2005; Williams et al., 2009). We intentionally chose not to change classroom practice dramatically, in order to gain greater buy-in on the part of the principals and the teachers. However, it appears that many teachers benefit from having very structured lessons prepared for them to deliver, and it is under these circumstances that intervention effects are most likely to be found (Guthrie, Wigfield, & Perencevich, 2004; Williams et al., 2005).

The type and amount of professional development that will help teachers successfully enhance children's comprehension of informational text may depend on the degree of change an intervention requires. Taking a broad focus, Guthrie, Wigfield, Barbosa et al., (2004) aimed at changing teachers' approach to reading instruction entirely. During a 12-week intervention, Guthrie et al. had teachers devote their daily 90-minute reading block to Concept-Oriented Reading Instruction (CORI). During CORI, they were to teach reading strategies in the context of hands-on science theme units that afforded children access to interesting texts and allowed them to make choices and collaborate. To make such a change possible, Guthrie et al. provided teachers with 10 days of professional development including instructional examples, materials, and time to plan lessons using a teacher's guide and with assistance available. In contrast, some interventions have had a narrower focus. Williams et al. (2009) focused on one aspect of informational text comprehension: compare-contrast text structure. They investigated the effect of teaching children to comprehend compare-contrast science texts in lessons that supplemented regular instruction. Teachers taught 45-minute sessions 3 times a week for a total of 22 sessions. Because teachers were supplied with detailed lesson plans and all necessary materials, professional development required only about 30 minutes.

The issue of how best to help teachers provide effective instruction relevant to informational text is an important area of investigation. Teachers are less comfortable with such text than with stories (Donovan & Smolkin, 2001; Yopp & Yopp, 2006), and the task of integrating informational text into literacy instruction is difficult for many. Ness (2011) noted that although teachers held a favorable view of informational text, they nevertheless reported under-using it due to factors such as lack of time, lack of resources, and curriculum constraints. Dreher and Zelinke (2010) reached similar conclusions based on a review of literature, but also identified teacher knowledge as a major challenge. Concerns cited in the literature include lack of knowledge about features and types of informational text, the availability of informational text, and the use of such text in instruction.

To be successful in implementing instruction, teachers need knowledge about informational text. Thus, both initial certification programs and professional development for practicing teachers should include more focus on the why and how of using this genre. However, even if teachers have extensive knowledge about informational text, they likely will need additional support to take full advantage of that knowledge. Neuman and Cunningham (2009) found that although a professional development course was effective in developing teacher knowledge, it did not result in improving teacher practice unless accompanied by coaching in which an expert helped a teacher learn to apply knowledge to practice.

Limitations in teacher knowledge and practice with respect to informational text may contribute to null effects on reading comprehension, not only in the REAL project but also in other recent interventions. In a large-scale study at fifth grade, James-Burdumy et al. (2010) found no effect on reading comprehension for supplemental reading curricula, all of which targeted reading to learn. Remarkably, these programs, with titles such as *Read for Real* and *Reading for Knowledge*, also did not lead to increased use of informational text. It seems likely that the lack of improvement in reading comprehension and the failure of teachers to increase the amount of informational text in their instruction are related. Although the nature of the instruction itself is of course important, not just an increased use of informational text (Purcell-Gates, Duke, & Martineau, 2007), teachers who have difficulty including more informational text may also have difficulty providing optimal instruction on such texts. Further research on what will assist teachers to increase their use of and facility with informational text may be particularly important if the Common Core State Standards effort is to be successful. The standards, in various stages of implementation across most states in the US, call for a 50-50 mix of literary and informational text in elementary school, both in grades K-2 and in grades 3-6 (Coleman & Pimentel, 2011a; 2011b).

Although there is much still to investigate, the REAL project provides an informative lesson for those who call for randomized controlled trials as the gold standard of educational research. In her 2005 presidential address to the Society for the Scientific Study of Reading, Joanna Williams referred to the study to illustrate the challenges of conducting classroom-based intervention research. Although the schools were located within the same district, each school had a different ethos and experienced several unique events each year of the study. One school received additional funding because students failed to show improvement on state assessments, and the principal used this money to implement small class sizes. In addition, the Text Infusion Alone school had a new principal in Year 3 who decided to departmentalize instruction in the fourth grade. Thus, all children in the study moved from classroom to classroom for language arts, science/social studies, and mathematics, and they were seldom given access to the informational books provided except during homeroom periods. Furthermore, the Traditional Instruction school became a magnet school for gifted and talented students during the course of the project, and many of the non-gifted students were transferred to other schools. And of course the small scale of the study meant that we were unable to control for teacher effects; teachers participating in the study varied dramatically in their experience and their effectiveness. The world of public schooling is such that it cannot be controlled.

• • •

Acknowledgements

The original research reported herein was funded by a grant from the Spencer Foundation to the first two authors. We gratefully acknowledge the foundation's support. Portions of this paper were presented at a symposium at the meeting of the Society for the Scientific Study of Reading in July 2005.

Linda Baker is Professor and Chair of the Department of Psychology at the University of Maryland, Baltimore County. Her research interests include motivation for reading, metacognition, prevention of reading difficulties, and parental influences on children's school outcomes. She is a former associate editor of the *Journal of Educational Psychology*, and she is a fellow of the American Psychological Association and the American Educational Research Association.

Mariam Jean Dreher is Professor of Reading Education at the University of Maryland, College Park. Her research interests include the effects of integrating informational texts into literacy instruction on children's comprehension, vocabulary, and motivation. She served as a Fulbright Senior Specialist working to help early childhood teachers in Finland explore those topics, and is currently a consultant to the National Geographic Society on a series of trade books for children.

Angela Katenkamp Shiplet received her Ph.D. in Applied Developmental Psychology and a certificate in American Educational Policy from the University of Maryland, Baltimore County in 2008. She served as the lead research assistant on the REAL project throughout its course. Her current interests include the relations between the school and the home.

Anita N. Voelker is an Associate Professor at Messiah College in Grantham, PA. She is currently researching how emergent readers respond to picture books on E-readers.

Lisa Carter Beall received her doctoral degree in Applied Developmental Psychology from the University of Maryland, Baltimore County where she studied developmental factors impacting children's learning. Her dissertation examined the role of implementation intentions and choice of reading material on the voluntary reading behavior of fourth-grade students. She is currently involved in assessment of first-year programs that are designed to facilitate the transition of freshman and transfer students to UMBC.

Adia J. Garrett, Ph.D. received her BS in Psychology from Morgan State University and a Ph.D. in Applied Developmental Psychology from the University of Maryland, Baltimore County (UMBC). She is currently a lecturer at UMBC who enjoys teaching courses in human development and research methods. Her research interest in the educational contexts of development and her commitment to helping students achieve success has been the foundation for her work in course redesign. She has collaborated with colleagues campus-wide to enhance pedagogy with technology, promote student engagement, and foster peer-assisted learning. Her most challenging, yet rewarding, job is being a new mom.

Heather R. Schugar is an Assistant Professor of literacy at West Chester University where she explores children's comprehension of informational texts and children's comprehension of interactive books available on e-Reader devices. She first worked on the REAL project as a doctoral student at the University of Maryland, College Park.

Maria Finger-Elam is a doctoral student in the Applied Developmental Psychology program at the University of Maryland, Baltimore County. Her dissertation focuses on parent involvement programs to help boost children's academic achievement. She is currently an evaluation specialist in the Howard County (Maryland) Public Schools Office of Research and Program Evaluation.

References

Aarnoutse, C., & Schellings, G. (2003). Learning reading strategies by triggering reading motivation. *Educational Studies, 29, 387-408.*

Baker, L. & Dreher, M. J. (2005, June). Balancing learning to read and reading for learning: Intervention effects on students' achievement. In M. J. Dreher & L. Baker (Chairs*), Does increasing the availability of informational text in urban classrooms improve young children's reading achievement and motivation?* Symposium presented at the meeting of the Society for the Scientific Study of Reading, Toronto.

Baker, L., Dreher, M. J., & Guthrie, J. T. (Eds.). (2000). *Engaging young readers: Promoting achievement and motivation*. New York: The Guilford Press.

Baker, L., & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly, 34,* 452-477.

Beall, L. C. (2011). *Implementation intentions as a means of promoting voluntary reading behavior* (Unpublished doctoral dissertation). University of Maryland, Baltimore County.

Beall, L. C., Morse, F., Baker, L., & Dreher, M. J. (2005, June). Student book preferences and their links to achievement. In M. J. Dreher & L. Baker (Chairs), *Does increasing the availability of informational text in urban classrooms improve young children's reading achievement and motivation?* Symposium presented at the meeting of the Society for the Scientific Study of Reading, Toronto.

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: IRA.

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29*, 137-164.

Campbell, J. R., Kapinus, B. A., & Beatty, A. S. (1995). *Interviewing children about their literacy experiences: Data from NAEP's Integrated Reading Performance Record (IRPR) at Grade 4*. Washington, DC: National Center for Educational Statistics.

Caswell, L. C., & Duke, N. (1998). Non-narrative as a catalyst for literacy development. *Language Arts*, *75*, 108-117.

Cervetti, G. N., Bravo, M. A., Hiebert, E. H., Pearson, P. D., Jaynes, C. A. (2009). Text genre and science content: Ease of reading, comprehension, and reader preference. *Reading Psychology, 30,* 487-511.

Chall, J. S., Jacobs, V. A., & Baldwin, E. (l990). The reading crisis: Why poor children fall behind. Cambridge, MA: Harvard University Press.

Chambliss, M. J., & McKillop, A. M. (2000). Creating a print- and technology-rich classroom library to entice children to read. In L. Baker, M. J. Dreher, & J. T. Guthrie (Eds.), *Engaging young readers: Promoting achievement and motivation* (pp. 94-118). Guilford Press.

Chapman, J. W., & Tunmer, W. E. (1997). A longitudinal study of beginning reading achievement and reading self-concept. *British Journal of Educational Psychology*, *67*, 279-291.

Chapman, M., Filipenko, M., McTavish, M., & Shapiro, J. (2007). First graders' preferences for narrative and/or information books and perceptions of other boys' and girls' book preferences. *Canadian Journal of Education 30*, 531-553.

Coleman, D., & Pimentel, S. (2011a). Publishers' criteria for the Common Core State Standards in English language arts and literacy, Grades K–2. Retrieved from http://www.corestandards.org /assets/Publishers_Criteria_for_K-2.pdf

Coleman, D., & Pimentel, S. (2011b). Publishers' criteria for the Common Core State Standards in English language arts and literacy, Grades 3–12. Retrieved from http://www.corestandards.org/ assets/Publishers_Criteria_for_3-12.pdf

Common Core State Standards (CCSS). (2010). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects*. Retrieved from http://www.corestandards.org/the-standards

Culatta, B., Hall-Kenyon, K. M., & Black, S. (2010). Teaching expository comprehension skills in early childhood classrooms. *Topics in Language Disorders, 30,* 323-338.

dePaola, T. (1984). *The popcorn book*. New York: Holiday House.

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

Donovan, C. A., & Smolkin, L. B. (2001). Genre and other factors influencing teachers' book selections for science instruction. *Reading Research Quarterly, 36*, 412–440.

Dreher, M. J. (2000). Fostering reading for learning. In L. Baker, M. J. Dreher, & J. T. Guthrie (Eds.), *Engaging young readers: Promoting achievement and motivation* (pp. 68-93). New York: Guilford Press.

Dreher, M. J., & Baker, L. (2005, June). Balancing learning to read and reading for learning: Infusing information books into primary-grade classrooms. In M. J. Dreher & L. Baker (Chairs), *Does increasing the availability of informational text in urban classrooms improve young children's reading achievement and motivation?* Symposium presented at the meeting of the Society for the Scientific Study of Reading, Toronto.

Dreher, M. J., & Voelker, A. (2004). Choosing informational books for primary-grade classrooms: The importance of balance and quality. In E. W. Saul (Ed.), *Crossing borders in literacy and science instruction: Perspectives on theory and practice* (pp. 260-276). Newark, DE: International Reading Association & Arlington, VA: The National Science Teachers Association.

Dreher, M. J., & Zelinke, S. B. (2010, December). *Informational text: Why is it still neglected in so many elementary classrooms?* Paper presented at the Literacy Research Association, Fort Worth, TX.

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., & Carlisle, J. (2011). The development of comprehension. In M. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research (*Vol. 4, pp. 199-228*)*. New York: Routledge.

Foorman, B. R., & Connor, C. M. (2011). Primary grade reading. In M. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research (*Vol. 4, pp. 136-156*)*. New York: Routledge.

Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading First Impact Study: Final Report (NCEE 2009-4038)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Guthrie, J. T., Hoa, A. L. W., Wigfield, A., Tonks, S. M., Humenick, N. M., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemporary Educational Psychology, 32*, 282–313.

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., 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: Erlbaum.

Hall, K. M., Sabey, B. L., & McClellan, M. (2005). Expository text comprehension: Helping primary-grade teachers use expository texts to full advantage. *Reading Psychology, 26*, 211-234.

Hoffman, J. V. (1992). Critical reading/thinking across the curriculum: Using I-charts to support learning. *Language Arts, 69*, 121-127.

Hoffman, J. V., McCarthey, S. J., Abbot, J., Christian, C., Corman, L., Curry, C. (1994). So what's new in the new basals? A focus on first grade. *Journal of Reading Behavior*, *26*, 47-73.

Horowitz, R., & Freeman, S. H. (1995). Robots versus spaceships: The role of discussion in kindergartners' and second graders' preferences for science texts. *The Reading Teacher, 49,* 30-40.

Ingham, J. (1982). *Books and reading development: The Bradford book flood experiment* (2nd ed.). Portsmouth, NH: Heinemann Educational Books Inc.

International Reading Association and National Council of Teachers of English (1996). *Standards for the English language arts*. Newark, DE: International Reading Association.

James-Burdumy, S., Deke, J., Lugo-Gil, J., Carey, N., Hershey, A., Gersten, R., Newman-Gonchar, R., Dimino, J., & Haymond, K. (2010). *Effectiveness of selected supplemental reading comprehension interventions: Findings from two student cohorts (NCEE 2010-4015)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education.

Jeong, J., Gaffney, J. S., & Choi, J.-O. (2010). Availability and use of informational texts in second-, third-, and fourth-grade classrooms. *Research in the Teaching of English, 44*, 435-456.

Katenkamp, A., Garrett, A., & Baker, L. (2005, June*).* Opportunities to read in the classroom: Observations of reading activities in Grades 2-4. In M. J. Dreher & L. Baker (Chairs), *Does increasing the availability of informational text in urban classrooms improve young children's reading achievement and motivation?* Symposium presented at the meeting of the Society for the Scientific Study of Reading, Toronto.

Kletzien, S. B., & Dreher, M. J. (2004). *Informational text in K-3 classrooms: Helping children read and write*. Newark, DE: International Reading Association.

Lewis, M., Wray, D., & Rospigliosi, P. (1994). " . . . And I want it in your own words." *The Reading Teacher, 47*, 528-536.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie Reading Test* (4th ed.). Itasca, IL: Riverside Publishing.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2002). *Gates-MacGinitie Reading Tests (4th ed.): Technical Report, Forms S and T*. Itasca, IL: Riverside Publishing.

McKenna, M. C., Ellsworth, R. A., & Kear, D. J. (1995). Children's attitudes towards reading: A national survey. *Reading Research Quarterly*, *30*, 934-956.

Mohr, K. A. J. (2003). Children's choices: A comparison of book preferences between Hispanic and non-Hispanic first-graders. *Reading Psychology, 24*, 163-176.

Moss, B. (2008). The information text gap: The mismatch between non-narrative text types in basal readers and 2009 NAEP recommended guidelines. *Journal of Literacy Research, 40*, 201-219.

Moss, B., & Newton, E. (2002). An examination of the informational text genre in recent basal readers. *Reading Psychology, 23*, 1-13.

Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy, P. (2007). *PIRLS 2006 international report: IEA's Progress in International Reading Literacy Study in primary schools in 40 countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

National Assessment Governing Board. (2008). *Reading framework for the 2009 National Assessment of Educational Progress*. Washington, DC: US Government Printing Office.

National Center for Education Statistics (NCES). (2009). *The Nation's Report Card: Reading 2009 (NCES 2010–458).* Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Council for the Social Studies (NCSS). (1994). *Expectations of excellence: Curriculum standards for social studies.* Washington, DC: Author.

National Council of Teachers of Mathematics (NCTM). Commission on Standards for School Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Research Council. (1994). *National science education standards: An enhanced sampler.* Washington, DC: National Academy Press.

Ness, M. (2011). Teachers' use of and attitudes toward informational text in K-5 classrooms. *Reading Psychology, 32*, 28-53.

Neuman, S. B. (2001). The role of knowledge in early literacy. *Reading Research Quarterly, 36*, 468-475.

Neuman, S. B. (2006). The knowledge gap: Implications for early education. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (pp. 29-40). New York: Guilford.

Neuman, S. B. (2010). Lessons from my mother: Reflections on the National Early Literacy Panel report. *Educational Researcher, 39*, 301-304.

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal, 46*, 532-566.

Oakhill, J. V., & Petrides, A. (2007). Sex differences in the effects of interest on boys' and girls' reading comprehension. *British Journal of Psychology, 98,* 223–235.

O'Hara, J. D. (2007). *The influence of supplemental instructional approaches upon the comprehension, metacognitive awareness, and motivation of struggling third- and fourth-grade readers* (Unpublished doctoral dissertation). University of Maryland, College Park.

Pappas, C. C. (1993). Is narrative "primary"? Some insights from kindergartners' pretend readings of stories and information books. *Journal of Reading Behavior, 25*, 97-129.

Purcell-Gates, V., Duke, N. K., & Martineau, J. A. (2007). Learning to read and write genre-specific text: Roles of authentic experiences and explicit teaching. *Reading Research Quarterly, 42*, 8-45.

Ruetschlin, H., Finger, M., & Dreher*, M. J. (2005, June)*. Children's reading comprehension in Grades 2-4 across genre and question type. In M. J. Dreher & L. Baker (Chairs*), Does increasing the availability of informational text in urban classrooms improve young children's reading achievement and motivation?* Symposium presented at the meeting of the Society for the Scientific Study of Reading, Toronto.

Salinger, T., Kamil, M. L., Kapinus, B., & Afflerbach, P. (2005). Development of a new framework for the NAEP reading assessment. *National Reading Conference Yearbook, 54*, 334–349.

Serpell, R. S., Baker, L., & Sonnenschein, S. (2005). *Becoming literate in the city: The Baltimore Early Childhood Project.* New York: Cambridge.

Smolkin, L., & Donovan, C. (2001). The contexts of comprehension: The information book read aloud, comprehension acquisition, and comprehension instruction in a first-grade classroom. *The Elementary School Journal, 102*, 97-122.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Taylor, B. M., & Pearson, P. D. (2001). *The CIERA School Change Classroom Observation Scheme*. University of Michigan: Center for the Improvement of Early Reading Achievement.

Venezky, R. L. (2000). The origins of the present-day chasm between adult literacy needs and school literacy instruction. *Scientific Studies of Reading*, *4*, 19-39.

White, S., Chen, J., & Forsyth, B. (2010). Reading-related literacy activities of American adults: Time spent, task types, and cognitive skills used. *Journal of Literacy Research, 42*, 276-307.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A., Freedman-Doan, C., & Blumenfeld, P. B. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A three-year study. *Journal of Educational Psychology*, *89*, 451-469.

Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology, 89*, 420-432.

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.

Williams, J. P., Stafford, K.B., Lauer, K. D., Hall, K. M., & Pollini, S. (2009). Embedding reading comprehension training in content-area instruction. *Journal of Educational Psychology, 10,* 1-20.

Yopp, R. H., & Yopp, H. K. (2006). Informational texts as read-alouds at school and home. *Journal of Literacy Research, 38*, 37-51.

This page is intentionally left blank

International Electronic Journal of Elementary Education, 2011, 4(1), 229-257.

Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge

Danielle S. McNAMARA*
Arizona State University, United States

Yasuhiro OZURU
University of Alaska Anchorage, United States

Randy G. FLOYD
The University of Memphis, United States

Abstract

We examined young readers' comprehension as a function of text genre (narrative, science), text cohesion (high, low), and readers' abilities (reading decoding skills and world knowledge). The overarching purpose of this study was to contribute to our understanding of the *fourth grade slump*. Children in grade 4 read four texts, including one high and one low cohesion text from each genre. Comprehension of each text was assessed with 12 multiple-choice questions and free and cued recall. Comprehension was enhanced by increased knowledge: high knowledge readers showed better comprehension than low knowledge readers and narratives were comprehended better than science texts. Interactions between readers' knowledge levels and text characteristics indicated that the children showed larger effects of knowledge for science than for narrative texts, and those with more knowledge better understood the low cohesion, narrative texts, showing a reverse cohesion effect. Decoding skill benefited comprehension, but effects of text genre and cohesion depended less on decoding skill than prior knowledge. Overall, the study indicates that the fourth grade slump is at least partially attributable to the emergence of complex dependencies between the nature of the text and the reader's prior knowledge. The results also suggested that simply adding cohesion cues, and not explanatory information, is not likely to be sufficient for young readers as an approach to improving comprehension of challenging texts.

Keywords: Comprehension, fourth grade slump, cohesion, genre, domain knowledge, reading, individual differences, coherence, construction integration

* Danielle S. McNamara, Department of Psychology Senior Research Scientist, Learning Sciences Institute Arizona State University, P.O. Box 872111, Tempe, Arizona 85287-2111, United States.
E-mail: Danielle.McNamara@asu.edu

ISSN:1307-9298

www.iejee.com

Introduction

A good deal of research has been conducted and has contributed to our understanding of how children learn to decode words and the factors that influence young readers' ability and inability to decode words (Cain, Oakhill, & Bryant, 2000; Ehri, 1991, Vellutino, Scanlon, & Spearing, 1995). Reading decoding represents the ability to apply letter-sound correspondence rules when reading words and non-words. Scholars have postulated that slow or inaccurate decoding skills tax working memory resources, using up working memory capacity needed for other comprehension processes such as integrating information across sentences (Cain, Oakhill, & Bryant, 2004; Hannon & Daneman, 2001; Perfetti, 1985). Indeed, slow or inaccurate word decoding has a profound impact on the reading comprehension success (Lyon, 2002; Vellutino, 2003).

There has also been a growing realization that children's ability to decode the words in text does not paint a complete picture of children's ability to comprehend text (e.g., Cain et al., 2004; Oakhill, Cain, & Yuill, 1998). The ability to decipher a word is not the same as the ability to interpret a sentence, understand the relationship between sentences, and to interpret the global meaning of a text (Oakhill, Cain, & Bryant, 2003). Successful comprehension also requires the reader to integrate individual word meanings into a coherent sentence level representation and to integrate sentences to create a global understanding. As such, successful reading comprehension requires the efficient coordination and integration of a number of underlying processes. These processes include not only word decoding and parsing sentences, but also integrating information within a text and with prior world knowledge (Kintsch, 1988, 1998; Perfetti, 1985).

More research now turns to developing a better understanding of children's comprehension processes (Cain et al., 2004; Cote & Goldman, 1999; Kendeou, van den Broek, White, & Lynch, 2009) as opposed to decoding processes. This project is intended to contribute to this understanding by examining the effects of both person-related and text-related factors on children's text comprehension. Thus, we examine two factors related to children's abilities (i.e., decoding skill and knowledge) and two factors related to text (i.e., cohesion and genre). Our research targets children in grade 4 because there is some evidence that children at that age are at a critical period in reading development characterized by an emergence of comprehension difficulties. This has been referred to as the *fourth grade slump* (Meichenbaum & Biemiller, 1998; Sweet & Snow, 2003). Our goal here is to examine the relative impact of the four factors related to children's ability and text characteristics so that we can more fully understand the potential problems leading to a fourth grade slump.

The guiding premise of this research is that world knowledge deficits, which are the negative gaps between the reader's actual knowledge and the knowledge demanded by a text to understand the text, are significant contributors to potential problems occurring when children reach the fourth grade. Thus, we examine here effects of knowledge on text comprehension, particularly in concert with characteristics of the text that influence the amount of knowledge required to understand the text. We further assume that genre and cohesion are two aspects of text that principally contribute to the degree of knowledge required to understand texts.

Based on these premises, we expect that readers generally understand (a) narrative text better than science text and (b) text with high cohesion better than text with low cohesion. In addition, we hypothesize the presence of an interaction between level of knowledge and text characteristics such that the benefit of knowledge is more pronounced for text with higher demands on knowledge. Thus, our specific predictions for the interactions are that the benefits of knowledge were expected to be more pronounced for science texts and for texts with more conceptual gaps (i.e., low cohesion texts). We explain these assumptions in greater detail in the following sections.

Text Comprehension

Our expectations are primarily based upon the Construction-Integration (CI) model of text comprehension (Kintsch, 1988, 1998). According to this theory, and indeed most theories of text comprehension (Graesser, Singer, & Trabasso, 1994; van den Broek, Rapp, & Kendeou, 2005), a critical process of successful comprehension is the retrieval of information from knowledge that is not explicitly stated in the text. According to the CI model, text comprehension has multiple levels, including a surface level representation of the words and syntax, and a textbase level that represents the meaning of the text. We hypothesize that, ultimately, the most important level of representation for comprehension that these children often struggle to construct is the situation model, which involves the integration of the textbase with knowledge. Comprehension is assumed to be more successful and deeper if the reader activates relevant knowledge and integrates that knowledge with the information explicitly stated in the text. In essence, text comprehension is more successful when the reader generates inferences while reading (Vidal-Abarca, Martinez, & Gilabert, 2000; Wolfe & Goldman, 2005).

Of course, successful comprehension is also largely dependent on the first two levels of comprehension. If the reader does not successfully form a surface-level representation, then the reader will be highly unlikely to form a coherent textbase. That is, if the reader does not decode the words or parse the sentences, the reader's surface level of comprehension will be deficient, and by consequence, the textbase will likely to be incoherent or malformed. If that is the case, then the activation of relevant knowledge and a coherent situation model representation are unlikely. In sum, the situational model generally builds upon the textbase and surface representations (unless, of course, the reader's understanding comes solely from prior knowledge and not from the text). Thus, the integration of knowledge with the textbase understanding requires sufficient decoding skill for a textbase to be formed. However, what is critical here is the notion that the contribution of decoding skill and knowledge work differently. That is, whereas decoding is fundamental to comprehension of texts across all genres with different features, the contribution of knowledge to comprehension is likely to vary depending on text genre and text features (in particular, text cohesion). Subsequent sections describe how knowledge contributes to comprehension depending on text genre and text cohesion.

Text Genre and World Knowledge

Our focus regarding the influence of text genre is on the distinction between narrative and expository texts, in particular science texts. As discussed in the previous section, world knowledge plays a critical role in deep-level comprehension of texts because readers must use knowledge to integrate meanings of individual sentences into a coherent representation of situations or events depicted by the overall text (Kintsch, 1988, 1998). As such, whether readers can develop a deep-level comprehension of the overall text meaning is likely to be affected by text genre. Narrative texts usually present reoccurring topics (e.g., friendship, love, and parting with a friend) in a specific context involving particular characters, settings, and times. Readers often have extensive experience and knowledge (i.e., schemas) regarding the events and situations described in typical narrative texts. Although narrative texts may contain new information (i.e., unfamiliar location, characters, and specific actions), most children have, from first-hand experience, well-developed schemas about the settings, actions and events described by narrative texts (Nelson, 1996; Olson, 1985). Thus, most children possess adequate event related knowledge to comprehend narrative texts. Moreover, many narrative texts also follow a simple structure—a sequence of casually related events for which many elementary school children are familiar (Williams et al., 2005).

In contrast to narrative texts, expository texts often place greater processing demands on the reader due to their increased structural complexity and increased demands for domain-specific information. Expository texts often contain abstract and logical relations that can be difficult to interpret, especially for children in the third to fifth grades (Kamberelis & Bovino, 1999). Perhaps most importantly, expository texts introduce many concepts that are new or only partially understood by the reader. Indeed, expository texts are used for the purpose of acquiring new information, and thus, they often contain novel content for young school children who are beginning to learn about those content domains, such as science. If children lack previous knowledge about a particular domain, comprehension will be limited because they do not possess the knowledge structures to which the new information can be integrated and assimilated (Langer, 1986).

The link between knowledge and expository text comprehension is well supported by previous research with adolescents and adults (Afflerbach, 1986; Chi, Feltovich, & Glaser, 1981; McNamara & Kintsch, 1996) and elementary school children (Best, Floyd, & McNamara, 2008; Rupley & Wilson, 1996). Thus, one possible interpretation for the emergence of comprehension difficulties around the fourth grade is that children lack sufficient prior knowledge to comprehend expository texts that are introduced during this period. Whereas early elementary school reading instruction focuses on the development of fundamental reading skills (i.e., learning to read), reading goals shift toward *reading to learn* in the third and fourth grades. Thus, up until the fourth grade, children tend to read narrative texts for the purpose of learning to read. However, as they transition from narrative text to expository texts to move to reading to learn during the third and fourth grades, and particularly in the fourth grade, knowledge levels may become the most critical influence on their comprehension.

Text Cohesion and World Knowledge

The effect of world knowledge on reading comprehension is also likely to be regulated by the manner in which reading materials are written. This issue is very important because both narrative and science texts can be written in different ways that might affect comprehension. The notion of text cohesion is one of the most useful concepts to systematically represent text characteristics that affect comprehension in a theoretically meaningful way. Text cohesion represents the extent to which a text explicitly provides background information and cues to help readers relate information distributed across different parts of the text (Britton & Gulgoz, 1991; Graesser, McNamara, & Louwerse, 2003). Cohesive elements in a text are grounded in explicit linguistic elements (i.e., words, features, cues, signals, constituents) and their combinations (Graesser & McNamara, 2010).

Texts are considered to be *low cohesion* when constructing a coherent representation from the text requires many inferences based on reader's knowledge. Texts are considered *high cohesion* when elements within the text provide more explicit clues to relations within and across sentences (McNamara, Louwerse, McCarthy, & Graesser, 2010). As such, low-cohesion texts place greater processing demands on the reader, in particular for readers with low levels of background knowledge. Previous research indicates that many expository materials written for school children have low levels of cohesion. For example, Beck, McKeown, and Gromoll (1989) performed an extensive analysis of four elementary school social studies texts and found that the texts comprised unclear goals and poor explanatory links and assumed too much knowledge on the part of readers. Thus, deficits in prior knowledge are likely compounded by exposure to low-cohesion texts.

In support of that hypothesis, a series of studies conducted by McNamara and colleagues (McNamara, 2001; McNamara & Kintsch, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996; O'Reilly & McNamara, 2007) indicates that the effects of domain knowledge in the comprehension of expository materials are moderated by text cohesion for middle school children and adults. Across these studies, the authors modified texts so that participants either read low-cohesion or high-cohesion versions of the same text. To form the high-cohesion texts, the low-cohesion texts were modified by adding surface-level indicators of relations between ideas in the text. Such modifications range from adding low-level information, such as identifying anaphoric referents, synonymous terms, connective ties, or headers, to supplying background information left unstated in the text (Beck, McKeown, Omason, & Pople, 1984; Beck, McKeown, Sinatra, & Loxterman, 1991; Britton & Gulgoz, 1991; for a review see McNamara et al., 2010). When consecutive sentences overlap conceptually, the reader is more likely to be successful in forming a coherent representation linking the meaning of the two or more sentences. Likewise, when relationships between ideas in the text are explicit by using connectives such as *because, consequently, therefore*, and *likewise*, the reader is more likely to understand the text content better.

These studies (McNamara, 2001; McNamara & Kintsch, 1996; McNamara et al., 1996; O'Reilly & McNamara, 2007) indicated that increased cohesion consistently facilitated comprehension for readers, in particular those with low levels of background knowledge. It was concluded that low-knowledge readers cannot easily fill in gaps in low-cohesion texts because they do not have the knowledge to generate the necessary inferences. Therefore, these readers need high-cohesion text to understand and remember the content. These studies also demonstrated a reverse cohesion effect, showing benefits from low-cohesion text for readers with high level of knowledge. Demonstrations of reverse cohesion supported the assumption that less cohesive texts force high-knowledge readers to generate knowledge-based inferences to bridge cohesion gaps present in the text, thus resulting in further integration of text information with pre-existing knowledge. We would like to emphasize that this gap-filling process can be successful only if readers have the sufficient amount of background knowledge that can be accessed or triggered based on limited textual information.

In this study, this gap-filling inference based on pre-existing knowledge would be most likely to occur for high knowledge readers' reading low cohesion narrative texts. It would be unlikely to occur for low-cohesion science texts because the level of background knowledge is still too low to afford such gap-filling inferences even among relatively high-knowledge students. We hypothesize that (most) grade 4 children will not have a sufficient knowledge base in science to automatically generate inferences when reading low cohesion science text. However, some grade 4 students are expected to have sufficient world knowledge relative to narrative texts. Thus, we expected that high-knowledge students would show a reverse cohesion effect for the narrative texts. In all other cases, we expected to find an advantage for the higher cohesion text.

Present Research

The overarching goal of this research is to further the understanding of the factors that lead to comprehension difficulties among elementary school children entering the period associated with the fourth grade slump. In light of this goal, we examine the roles of reading decoding skill and world knowledge among children in the fourth grade when exposed to texts from different genres (narrative and expository) and different levels of cohesion.

Our study examined elementary-school children's comprehension of narrative and expository texts used in the classroom. We examined separately the effects of knowledge and word decoding skill, and how the effects of text characteristics (i.e., genre and cohesion) depend on knowledge or decoding skill. We expected to find significant effects of both knowledge and decoding skill on comprehension across different texts. We further expected that children would encounter greater difficulty comprehending the science texts and low cohesion texts, both of which are more knowledge demanding.

Most importantly, we hypothesized that comprehension would depend on both knowledge and the characteristics of the text. We predicted that comprehension of science texts, in contrast to narrative text, would depend on world knowledge. In contrast, we did not expect such an interaction between children's decoding skills and text types. Specifically, an interaction was expected between text genre and the readers' level of knowledge, wherein knowledge has a greater effect on expository text comprehension than narrative text comprehension. Decoding skill was expected to benefit comprehension, but a differential effect of decoding skill as a function of text genre was not predicted.

We also hypothesized that the effects of text cohesion would depend on both knowledge and text genre. Specifically, we predicted an advantage for low cohesion texts (i.e., a reverse cohesion effect) when the texts were relatively familiar (i.e., narratives) and when the reader had sufficient knowledge to fill in the cohesion gaps (i.e., high knowledge readers). In all other cases, we expected to find an advantage for higher cohesion text.

Method

Participants

Participants included 65 children enrolled in the fourth grade at four public schools in a large metropolitan school district. Children ranged in age from 9 years, 2 months to 11 years, 2 months (M = 118.30 months, SD = 5.35 months). Girls composed 52.3% of the sample (n = 34), and boys composed 47.7% (n = 31). Of the sample, 54% were Caucasian (n = 35), 40% of the children were African American or Black (n = 26), and 3% were Hispanic (n = 2). All children but two spoke English as their primary language. Using parent education level as an index of socioeconomic status, 2% of fathers did not complete high school; 49% of mothers and 55% of fathers graduated from high school, completed some college, or completed technical school; and 49% of mothers and 35% of fathers obtained at least a college degree.

On two screening measures, children in this sample demonstrated vocabulary knowledge and listening comprehension skills that were somewhat above average for their age. The average performance on the Woodcock–Johnson III (WJ III) Tests of Achievement (ACH) Picture Vocabulary test (Woodcock, McGrew, & Mather, 2001) was 108.2 (SD = 10.7), and average performance on the WJ III ACH Oral Comprehension test was 108.0 (SD = 9.5). Because the population means and standard deviations for these tests are 100 and 15, respectively, these results also indicate that the participating children displayed, on average, somewhat less variability than expected of the population of children this age.

Design

The experimental design of the study was a 2 x 2 within-subjects design. The within-subjects factors were text genre (narrative and expository) and text cohesion (low cohesion and high cohesion).

Materials

Texts. There were eight texts used in this study. Four texts were original texts obtained from basal readers and science textbooks. These texts were considered low-cohesion texts. The four remaining texts were revised versions of the four low-cohesion texts that were manipulated to increase their cohesion (see Appendix 1). The four low-cohesion texts included two science and two narrative texts. They were drawn from a pool of 127 texts collected from elementary-school-age basal readers and science textbooks. These texts included 67 science texts, 53 narrative texts, and 7 science texts written in narrative format. The average number of words per text was 388 (*SD* = 167, Min = 115, Max = 991; Science = 388; Narrative = 399; Mixed Format = 298), and the average Flesch-Kincaid reading level was 3.86 (*SD* = 1.79, Min = 0, Max = 7.8; Science = 4.96; Narrative = 2.52; Mixed format = 3.53). From this pool of 127 texts, we selected two narrative texts and two expository texts. These four texts were chosen because they were representative of the text pool and closely equated in terms of number of words and Flesch–Kincaid grade level. These indices were derived using Coh-Metrix, Version 1.0 (Graesser, McNamara, & Louwerse, & Cai, 2004). Preliminary selection criteria for inclusion in this study were for Flesch–Kincaid grade level to be between 2.0 and 5.0 and text length to be within the range of 304 and 471 words. From the text pool, we selected two science texts, *Heat* (SRA's *Real Science, Grade 2: Elementary Science*) and *Needs of Plants* (McGraw–Hill's *Science, Grade 2*), and two narrative texts, *Moving* (McGraw-Hill *Reading, Grade 3*) and *Orlando* (Addison Wesley's *Phonics Take-Home Reader, Grade 2*).

The cohesion of the four original texts was then manipulated to increase their cohesion. Each of the two texts within each genre included one original, lower-cohesion version and one higher-cohesion version. The aim of the cohesion manipulations for the high-cohesion versions was to increase cohesion between concepts and ideas such that they created a clear situation model for the child. The basic concepts discussed in the low-cohesion and high-cohesion texts were the same (e.g., heat moves through objects). However, the understandability of the high-cohesion versions increased by cohesion cues. There were a number of cohesion manipulations made to the high-cohesion texts that were fitting for each of the four texts. We increased cohesion using methods previously found to enhance comprehension (e.g., Beck et al., 1991; McNamara et al., 1996), including manipulations to referential, temporal, causal and explanatory cohesion. Specifically, there were seven aspects of the text that were modified to increase cohesion: (a) replacing pronouns with noun phrases, (b) adding descriptive elaborations, (c) adding sentence connectives, (d) replacing or inserting words to increase conceptual overlap, (e) adding topic headers, (f) adding theme sentences, and (g) moving or re-arranging sentences to increase temporal or referential cohesion. For example, if events were not presented in chronological order, the presentation of the events was altered to match the chronological order of the events in the world. The aim was to alter the texts so that they approximated equivalent levels of cohesion as measured by the Coh-Metrix, Version 1.1 (Graesser et al., 2004). The tool automatically analyzes texts on over 50 types of cohesion relations and over 200 measures of language and discourse by applying modules that use lexicons, classifiers, syntactic parsers, shallow semantic interpreters, conceptual templates, latent semantic analysis, and other components widely used in computational linguistics.

The high-cohesion versions of the texts included explicit information about the meanings of particular terms, a greater number of noun phrases, and a greater number of causal connections. The following example taken from the plant texts illustrate the ways in which cohesion was added. In this example, the high-cohesion version adds a sentence explaining that a mineral is not a plant or an animal. The third sentence includes a connective term "instead."

Low cohesion. "Plants also need minerals. A mineral is a naturally occurring substance that is neither plant nor animal."

High cohesion. "Plants also need minerals. A mineral is not a plant or an animal. Instead, a mineral is a substance in the ground that occurs naturally."

Cohesion manipulations also involved creating a context so that the child could more easily interpret the situations described in the text. The following example taken from the opening sections of the Orlando texts illustrates an instance in which a context was created for the high-cohesion version of the Orlando text. It is also important to point out that the order in which information was presented was changed for the Orlando text such that the high-cohesion version provided greater temporal cohesion, that is, information was presented in the order in which events occurred; the low-cohesion version on the other hand presented information in a non-temporal order and thus the reader was to infer that information presented at the start of the story was not the first event to occur.

Low cohesion. "Salvador was upset. He told is Mama he was going out. He didn't want to be worried or sad."

High cohesion. "Once upon a time there was a boy. His name was Salvador. Salvador adored his pet pig named Orlando."

A further method for increasing text cohesion was to integrate information across sentences in the low-cohesion text to provide a clearer depiction of the situations described by the text. The following example taken from the Heat texts illustrates an instance in which information was integrated.

Low cohesion. "Most metals are good conductors. Metal pots are used for cooking. Heat from the stove quickly moves through the metal. The heat warms the food."

High cohesion. "Most metals are good conductors. For example, metal pots are used for cooking because heat from the stove quickly moves through the metal pots and the heat from the pot warms the food."

Table 1. *Select Characteristics of the Narrative and Expository Texts*

	Narrative				Expository			
	Orlando		Moving		Plant		Heat	
	Low	High	Low	High	Low	High	Low	High
Argument overlap	0.35	0.62	0.53	0.85	0.78	0.86	0.66	0.84
Number of words	451	547	437	584	466	637	404	521
Average sentence length	6.44	9.77	12.49	14.60	10.13	10.98	7.21	11.33
Grade level	2.20	3.86	4.02	4.82	3.76	3.77	2.67	4.93
Word frequency (min log)	1.85	1.54	1.86	1.83	1.46	1.48	1.50	1.40

Table 1 presents some of the main text characteristics for the eight texts. These features are argument overlap, number of words, average sentence length (i.e., average number of words in a sentence), Flesch-Kincaid Grade Level, and average of word frequency of the lowest word frequency word in each sentence (logarithm). The argument overlap scores, which relate to the proportion of adjacent sentences that share one or more arguments (e.g., pronoun, noun or noun phrase), were higher for the high-cohesion versions of the texts. Incorporating words and phrases to increase cohesion has an effect on text length, such that the high-cohesion versions of the texts comprised more words than the low-cohesion versions. Also, adding cohesion changed the number of sentences and sentence length as indicated in the table. In addition, manipulating texts for cohesion affected the grade level scores, such that the high-cohesion versions were estimated as having higher grade-level scores than low-cohesion versions. Grade level assignments are primarily based on factors such as the number of words in the sentences and the number of letters or syllables per word (i.e., as a reflection of word frequency). Thus, adding words and sentence length (which is generally necessary to increase cohesion), increases grade level scores. However, the cohesion manipulation did not greatly influence content-based difficulty as indicated by word frequency.

Comprehension Measures.

Comprehension was assessed using a combination of recall tasks and multiple-choice questions. Multiple measures have the benefit of providing a more thorough evaluation of the breadth and depth of comprehension.

We collected free recall and cued recall data to assess children's comprehension of the texts. The free recall task required children to report what they remembered about the passage they had just read. Each child was provided the following directive: "Tell me everything you can remember about what you have just read. Give me as many details as possible, like you were trying to tell a friend about what you just read." All responses were recorded on an audiotape and later transcribed.

Cued recall, which assessed major themes in the text, was used to evoke richer content from the children (Zinar, 1990). The cued recall task required children to respond to three directives. The directives were designed to assess comprehension of three major sections of the texts, and they essentially covered the entire text. For example, for the Orlando text, children were directed to (a) "Tell me everything that Salvador did after his mama told him they would have to sell Orlando," (b) Tell me everything that Salvador's mama said at first about what they needed to do with Orlando," and (c) "Tell me everything that happened to Salvador and what he did after the storm began." All responses were audiotaped and later transcribed.

Twelve multiple-choice questions were constructed for each text to assess students' comprehension. Six of the questions were designed to tap local-level comprehension, and the other six were designed to tap global-level comprehension of the text. Whereas local-level questions requested information that was within five or fewer clauses (mostly within 2 sentences), global-level questions requested information that was located across six or more clauses. Each multiple-choice question had four answer options with only one being the correct answer. Examiners read the questions orally while the questions were presented in text form. The children were required to vocalize the correct answer. Comprehension scores for each text were obtained for each child by calculating the proportion of correct responses to total questions (i.e., 12).

Reading Competency Measures

As part of a larger battery of assessment instruments, children completed two tests from the WJ III ACH (Woodcock et al. 2001). The tests from the WJ III ACH included the Word Attack test and the Academic Knowledge test. Word Attack measures reading decoding skills. Examinees must pronounce phonologically regular non-words. The test has a median internal consistency reliability coefficient of .94 for ages 8 to 10 (McGrew & Woodcock, 2001). Academic Knowledge measures knowledge about the biological, physical, and social sciences and the humanities. Examinees must provide information about the biological and physical sciences; about history, geography, government, and economics; and about art, music, and literature. The test has a median internal consistency reliability coefficient of .84 for ages 8 to 10 (McGrew & Woodcock, 2001). For both WJ III tests, age-based standard scores

(*M* = 100, *SD* = 15) were obtained. For Word Attack, the standard score represents decoding skills. The standard score for the Academic Knowledge tests represents world knowledge.

Procedures

Recruitment. Children were recruited by sending letters of invitation to parents through the children's school classrooms. The letters provided information about the study and requested that parents contact researchers to schedule a testing session. Testing sessions were conducted on five Saturdays during fall 2003. After completion of the testing, children were provided a gift card to a department store, coupons from merchants, and school supplies.

Testing. An assessment battery was completed in approximately a 2-hour testing session. Graduate students who had successfully completed a graduate course covering the administration of standardized tests completed all testing. Children first silently read each text within a 5-minute period. After reading the first text, the text was removed from view before answering the free and cued recall questions and 12 multiple-choice questions. This process was repeated with the remaining four texts. The cohesion manipulation was organized such that children either read the high-cohesion or low-cohesion version of each text. The cohesion manipulation was counterbalanced so that an equal number of children read high-cohesion and low-cohesion versions of each of the four texts. The order of texts was counterbalanced using a Latin-square design. Finally, after reading the texts and answering the recall and multiple-choice questions, children completed the battery of reading competency tests.

Coding Recall Data. The analysis focused on the amount of information children recalled about information in the text by counting the number of propositions recalled for each text. It is important to note that we matched all recall to ideas contained in the low-cohesion versions of the texts so we could evaluate whether reading the high cohesion versions of a texts that contained the same ideas in a more understandable way, increased recall for the main ideas contained in the low-cohesion versions of the text. Thus, the number of propositions in the low-cohesion texts provided a benchmark for which we could compare the amount of recall that children generated from the low-cohesion and high-cohesion versions of the texts.

There were two steps to the recall analysis. First, the low-cohesion versions of narrative and expository texts were propositionalized using a conventional method in which the information contained in each sentence was broken into main propositions and sub-propositions (see Kintsch, 1998). Main propositions consisted of the main idea, whereas the sub-propositions contained details pertaining to the main idea. For instance, the main proposition for the sentence "Plants need sunlight, air and water to live" consisted of the notion that plants need things to live. The sub-proposition consisted of the notion that plants need sunlight, and water. The number of main propositions mapped on to the number of sentences contained in each low-cohesion text (e.g., 45 propositions for the plant text). Second, the children's transcribed recall data were divided into idea units. Idea units were classified as utterances that contained a subject, verb, and direct object. Idea units were separated by connectives, such as so, and, but, and because. Every idea unit was matched to the propositions. In cases where children repeated information, each idea unit was counted only once.

Our initial analysis also focused on inferences children generated about information in the text (i.e., information that was extrapolated from but not directly specified in the texts). Using previous research as a guide (e.g., Kintsch, 1993), inferences were classified as text-based, elaborative, global, or irrelevant. However, we did not further analyze inference data because too few inferences were produced (inferences comprised only 3% of cued recall question answers and 1% of free-recall question answers).

All the free recall and cued recall was coded by two trained raters. Half the data was coded by a third trained rater. Inter-rater reliability was evaluated for all dimensions of coding between the third rater and each of the two raters. Simple agreement and Kappa analyses indicated that agreement reached 90% or above on all dimensions, which indicates a high level of agreement. Disagreements were resolved by discussion between raters.

Proposition Analysis. The recall analyses assessed the number of propositions recalled. To account for the completeness of information recalled, a value of 1.0 was assigned to recall that contained the main proposition and more detailed information cited in the sub-proposition. A value of 0.5 was assigned to recall data that contained the main proposition but that did contain the detailed information in the sub-proposition. A value of 0 was assigned when the proposition was not recalled. For example, a value of 0.5 was assigned for the sentence "Plants need sunlight, water, and air to live" when a child stated that plants need water to live. A value of 1 was assigned when the child stated that plants need water, sunlight, and air. Because children's propositional recall sometimes contained erroneous information, recall that contained such information received a score of 0.5. For example, a child may have erroneously stated, "Plants do not need sunlight." Once recall scores were totaled for a text, they were transformed into a proportion because there were an unequal number of propositions contained in the four texts.

Free recall and cued recall propositions were summed into separate scores to focus on different dimensions of comprehension across tasks. Thus, it was possible for children to recall the same information in the free recall task and in the cued recall task. For the free recall analysis, proportion scores were calculated by dividing the free recall score by the number of propositions that could potentially be recalled for each text. For the cued recall analysis, the recall scores were summed for each of the free recall questions and divided by the number of directly relevant main propositions that could be potentially recalled (for each question).

To determine which sentences contained directly relevant information, two experimenters identified propositions that directly related to the cued-recall directives from three categories: directly related sentences, indirectly related sentences, and irrelevant sentences. There were 32 main propositions directly relevant to the Plant directives, 52 to the Heat directives, 28 to the Moving directives, and 35 to the Orlando directives. Note that the same proposition was sometimes classified as directly relevant for more than one cued-recall question; in such cases, the same proposition was counted for each of the relevant questions. Kappa analyses showed that there was a high level of agreement between raters across both texts (weighted Kappa = .85 expository texts and .85 for the narrative texts). Disagreements were resolved by discussion.

Results

Preliminary Analyses

Performance on the comprehension questions for the two narrative texts (Orlando and Moving) and for the two expository texts (Plants and Heat) were compared to verify whether there was an effect of text within genre. There were no statistically significant differences in children's performance comparing the two narrative texts on the multiple-choice, free recall, or cued recall tasks. There were no differences on children's performance between the two expository texts for the free recall and cued recall tasks. However, on the multiple-choice questions, children performed better on the Plant questions (M = .60, SD = .22) than on the

Heat questions (*M* = .52, *SD* = .21), *t*(64) = 3.96, *p* < .01. We conducted item analyses to examine whether these effects were due to a subset of the multiple-choice questions, but superior performance was evident for the majority of questions for the Plants text. We further confirmed that this effect may have been due to the ease of the questions (and not due to differences between the texts) because the Flesch–Kincaid grade level for the multiple-choice questions for the Heat text (*M* = 2.99, *SD* = 1.77) was somewhat, though not significantly, higher than the Plant text (*M* = 1.94, *SD* = 1.29), *F*(1, 22) = 2.751, *p* = .111, indicating that questions for the Heat text are somewhat more difficult than questions for the Plant text. These effects are unfortunate; however, the results from the recall analyses were not affected by the ease of the questions and thus provide confirmatory validity to the overall results.

Relations between Comprehension Measures, Decoding Skill, and World Knowledge

As can be expected, there was a significant, moderate correlation between decoding skill and world knowledge (*r* = .52, *p* <.01). Among dependent measures, the two recall measures (free and cued) correlated highly (*r* = .81, *p* <.01). By contrast, where the tasks differed more, the correlations between performance on the multiple-choice questions and free recall (*r* = .46, *p* <.01) and cued recall (*r* = .50, *p*<.01) were significant but moderate in magnitude.

Effects of Reader Abilities, Genre, and Cohesion

We performed a median split on the standardized individual difference scores resulting in high and low groups for each knowledge and decoding skill measure. The mean scores for the high and low groups on the two individual difference measures are presented in Table 2.

Table 2. *Mean Scores on Aptitude Measures for Children Assigned to the Low and High Groups*

Individual difference test	Ability group	
	Low *M* (*SD*)	High *M* (*SD*)
WJ III Academic Knowledge	92.97 (5.04)	113.06 (6.80)
WJ III Word Attack	97.39 (4.04)	114.17 (7.86)

We conducted separate mixed ANOVAs on performance on each of the three dependent measures: multiple-choice questions, free recall questions, and cued recall questions. In these ANOVAs, within-subjects factors were genre (narrative vs. expository) and cohesion (low vs. high cohesion). The between-subjects factor was either knowledge level (high vs. low knowledge) or reading decoding skill (high vs. low decoding skill). The ANOVAs were conducted separately for knowledge and decoding skill because including both factors reduced some cell sizes to unacceptably low numbers.

To ensure that the statistically significant effects identified from the ANOVAs were not distorted by entering knowledge or decoding skill as a categorical variable (high or low), we conducted multiple regression analyses in which either knowledge and reading decoding skill was entered as a continuous variable. Children's scores on the Academic Knowledge test or scores on the Word Attack test were entered as a separate independent variable, and scores on the comprehension measures were the dependent variables. We also ran multiple regression analysis with the Academic Knowledge and Word Attack scores as well as the centered interaction between the Academic Knowledge and Word Attack scores as independent variables to assess the robustness of these effects comprehension. Significant main and interaction effects generated from the ANOVAs were replicated in the regression analyses. We report only the ANOVAs because this method mimics prior research on knowledge and cohesion interactions, and is most easily interpreted.

World Knowledge Analysis

We begin by describing the results when knowledge levels are entered as the between-subjects factor. A 2 x 2 x 2 mixed ANOVA was conducted for each of the three dependent measures. These data are presented in Table 3. The first section below describes the effects of genre and knowledge, and the second describes the effects of cohesion.

Table 3. *Proportion Correct for the Three Reading Comprehension Measures as a Function of Knowledge, Genre, and Cohesion*

Knowledge level	Text type	Reading comprehension measure		
	Genre/Cohesion	Multiple-choice	Cued recall	Free recall
		M (*SD*)	*M* (*SD*)	*M* (*SD*)
Low	Narrative/Low	.62 (.16)	.33 (.15)	.16 (.12)
Low	Narrative/High	.70 (.12)	.36 (.15)	.18 (.11)
Low	Expository/Low	.45 (.17)	.06 (.05)	.06 (.05)
Low	Expository/High	.42 (.18)	.06 (.05)	.04 (.03)
High	Narrative/Low	.77 (.14)	.39 (.14)	.22 (.11)
High	Narrative/High	.81 (.11)	.30 (.12)	.19 (.09)
High	Expository/Low	.68 (.17)	.12 (.08)	.08 (.06)
High	Expository/High	.71 (.17)	.15 (.09)	.09 (.07)

Table 4. *Proportion Correct for Three Reading Comprehension Measures for the Narrative and Science Texts*

	Narrative text	Science text	ANOVA	
	M (*SD*)	*M* (*SD*)	*df1, df2*	*F*
Fourth-grade children				
Multiple-choice	.73 (.12)	.57 (20)	1,63	68.20***
Cued recall	.34 (.12)	.10 (.07)	1,50	226.72***
Free recall	.19 (.09)	.07 (.05)	1,52	113.92***

Note. *** = $p < .001$

Table 5. *Proportion Correct for Three Reading Comprehension Measures for Low and High-Knowledge Readers*

	Low Knowledge	High Knowledge	ANOVA	
	M (*SD*)	*M* (*SD*)	*df1, df2*	*F*
Multiple-choice	.34 (.10)	.74 (.10)	1,63	59.27***
Cued recall	.20 (.07)	.24 (.07)	1,50	2.88
Free recall	.11 (.06)	.15 (.07)	1,52	3.68

Note. * = $p < .05$, *** = $p < .001$.

Table 6. *Interaction between Genre and Knowledge*

	Narrative			Expository			
	Low K *M* (*SD*)	High K *M* (*SD*)	*t*	Low K *M* (*SD*)	High K *M* (*SD*)	*t*	Interaction *F*
Multiple-choice	.66 (.12)	.79 (.87)	-5.20***	.43 (.14)	.70 (.15)	-7.36***	11.06**
Cued recall	.34 (.13)	.34 (.11)	-0.50	.04 (.03)	.09 (.05)	-5.21***	2.99
Free recall	.17 (.10)	.20 (.08)	-1.22	.04 (.04)	.08 (.05)	-2.85**	<1

Note. K= knowledge.
** = *p* < .01, *** = *p* < .001.

Genre and knowledge. As shown at the top of Table 4, there was a main effect of genre for all three dependent measures, indicating that children's comprehension was better for the narrative texts than for the expository texts. As shown at the top of Table 5, the effect of knowledge was significant for the multiple-choice questions but marginal for the cued recall and recall measures. Thus, high-knowledge readers showed better comprehension than did low-knowledge readers, and this finding was most apparent on the multiple-choice questions.

As shown in Table 6, there was a significant interaction between genre and knowledge using scores from the multiple-choice questions, and there was a marginal interaction using the cued recall measure. Follow-up tests confirmed that the effect of knowledge was greater for the comprehension of expository texts than for the narrative texts.

Text cohesion. There was a main effect of cohesion on the multiple-choice question performance. As shown in Table 7, children comprehended the high-cohesion texts better than the low-cohesion texts. However, the effects of cohesion were not reliable for the two recall measures.

Table 7. *Proportion Correct for Low and High Cohesion Texts on the Three Dependent Measures*

	Low cohesion *M* (*SD*)	High cohesion *M* (*SD*)	*F*
Multiple-choice	.63 (.15)	.67 (.15)	4.35*
Cued recall	.22 (.09)	.22 (.08)	< 1
Free recall	.13 (.7)	.12 (.7)	< 1

Note. * = *p* < .05.

Table 8. *Interaction between Cohesion and Genre*

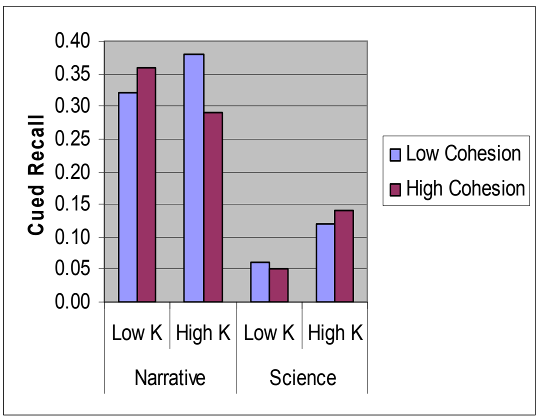
	Narrative			Expository			
	Low Coh *M* (*SD*)	High Coh *M* (*SD*)	*t*	Low Coh *M* (*SD*)	High Coh *M* (*SD*)	*t*	Inter-action *F*
Multiple-choice	0.69 (0.16)	0.76 (0.13)	3.09**	0.57 (0.21)	0.57 (0.23)	0.12	5.07*
Cued recall	0.35 (0.15)	0.33 (0.14)	1.36	0.09 (0.06)	0.10 (0.09)	-1.26	5.66*

Note. Coh = cohesion.
** = *p* < .01, *** = *p* < .001.

As shown in Table 8, there was a significant interaction between cohesion and genre on the multiple-choice question and cued recall question performance. Follow-up analyses on the multiple-choice data indicated that for the narrative texts, children comprehended the high-cohesion texts better than the low-cohesion texts, whereas there was no effect of cohesion for the expository texts. The interaction according to the cued recall measure is more difficult to interpret because none of the separate effects were significant. Nonetheless, the trends indicated that there was a slight advantage for low cohesion narrative texts, compared to a slight advantage for high cohesion science texts.

There was also a two-way interaction between cohesion and world knowledge on the cued recall task, $F(1,50) = 5.67$, $p < .05$, but not on the multiple-choice or free recall task (both $F<1$). The interaction indicated that high-knowledge readers recalled more information after reading the low-cohesion texts, $M = 0.26$, $SD = 0.09$, than the high-cohesion texts, $M = 0.22$, $SD = 0.08$; $t(25) = -2.23$, $p < .05$). However, there was no effect of cohesion for low-knowledge readers, $M_{low\ cohesion} = 0.20$, $SD = 0.08$; $M_{high\ cohesion}=0.21$, $SD = 0.08$; $t(25) = 1.10$, $p = .28$.

Most importantly, there was a three-way interaction between genre, cohesion, and knowledge on the cued and free recall measures (cued recall: $F(1,50) = 11.86$, $p < .01$; free recall; $F(1,52) = 6.52$, $p <.05$). This interaction was marginal for the multiple-choice measure, $F(1,63) = 2.61$, $p = .10$. This interaction is displayed for the recall measures in Figure 1 (see also Table 3 for means). As already noted, there is a larger effect of knowledge for the science texts than for the narrative texts. Moreover, we see in Figure 1 that the effect of knowledge depends on cohesion, particularly for the narrative texts, such that there was a reversed cohesion effect for high-knowledge readers and a slight benefit for high cohesion for the low-knowledge readers. High-knowledge students tended to show better comprehension after reading the low-cohesion narrative text than the high cohesion narrative text (multiple choice, $t(33) = -1.51$, $p = .14$, cued recall, $t(26) = -3.88$, $p < .05$, free recall, $t(26) = 1.65$, $p = .11$). In contrast, this pattern did not emerge for the low-knowledge readers. There were trends indicating benefits of cohesion for low-knowledge readers on the narrative text and high-knowledge readers on the science texts, but these effects were not reliable. It is notable that the scores for low-knowledge readers on the measures of expository recall were extremely low (see Figure 1 and Table 3), indicating that low-knowledge readers may not have understood the science text sufficiently in order for us to fully interpret the three-way interaction.



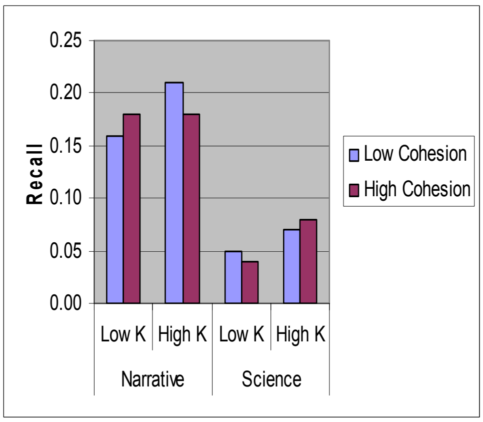


Figure 1. Interaction between knowledge, genre, and cohesion on the cued recall task and the free recall task.

Decoding Skill Analysis

Our second set of analyses examined potential effects of decoding skill for the fourth grade children by conducting the same 2 x 2 x 2 ANOVA as reported above, but this time with decoding skill entered as the dichotomous variable. We only report the effects that are not redundant with the effects reported above, that is, those that involve decoding skills. The mean scores on the comprehension measures are reported in Table 9.

Table 9. *Proportion Correct Scores as a Function of Decoding Skills, Genre, and Cohesion*

Decoding skill	Genre/cohesion	Multiple-choice *M* (*SD*)	Cued recall *M* (*SD*)	Free recall *M* (*SD*)
Low	Narrative low	.65 (.17)	.32 (.14)	.17 (.10)
Low	Narrative high	.72 (.13)	31 (.12)	.18 (.10)
Low	Expository low	.50 (.19)	.08 (.06)	.06 (.04)
Low	Expository high	.51 (.23)	.10 (.07)	.05 (.03)
High	Narrative low	.75 (.15)	.41 (.15)	.22 (.13)
High	Narrative high	.81 (.11)	.35 (.16)	.21 (.11)
High	Expository low	.66 (.20)	.11 (.08)	.09 (.06)
High	Expository high	.65 (.20)	.11 (.10)	.10 (.08)

Table 10. *Proportion Correct for Low and High Decoding Children on the Three Dependent Measures*

	Low decoding *M* (*SD*)	High decoding *M* (*SD*)	*F*
Multiple-choice	.59 (.14)	.72 (.11)	15.54***
Cued recall	.20 (.07)	.25 (.08)	4.28*
Free recall	.11 (.05)	.15 (.07)	5.05*

Note. * = $p < .05$, *** = $p < .001$.

As shown in Table 10, there was a main effect of decoding skill according to all three measures of comprehension indicating that skilled decoders performed better on the comprehension tasks than their less skilled counterparts. However, there were no significant interactions involving decoding skills. Thus, decoding skill affected comprehension, but the effects of text genre and text cohesion did not depend on decoding skill.

Discussion

The purpose of this study was to examine young children's comprehension as a function of the characteristics of the text and the children's knowledge and reading decoding skill levels. Our overarching goal was to more fully understand factors that might contribute to reading problems that seem to emerge in the fourth grade, called the fourth grade slump. The most important prediction made in this study is that at the age when young children are expected to begin *learning from text*, successful comprehension will largely depend on the reader's knowledge about the world and about specific domains. We made this prediction because this educational period is when children are exposed to a wider variety of texts, and moreover, successful comprehension of those texts demands the integration of knowledge with the text. The more that knowledge is required from the text, the more we can expect to see influences of the reader's knowledge level. Thus, we predicted that interactions between levels of world knowledge and characteristics of the text would emerge around grade 4. Indeed, we observed complex interactions between knowledge, text genre, and text cohesion in this study with fourth grade students.

Findings confirmed that text comprehension is influenced by decoding skill as well as the availability of knowledge to the reader. Better decoders showed better comprehension than less skilled decoders and children with more world knowledge performed better on the comprehension measures than did those with less knowledge. Also, children showed better comprehension of narrative than science texts. This result is most likely because narrative texts are more familiar in structure and contain more familiar information than do science texts, and thus readers have more knowledge available to facilitate comprehension.

We also found that the effect of knowledge was greater for science texts than for narrative texts. Thus, knowledge was more important when the text was more knowledge demanding. Further, a three-way interaction was expected between knowledge, text genre, and text cohesion. The children exhibited a reversed cohesion effect when the texts were familiar (i.e., narrative texts) and when they had sufficient knowledge to fill in the cohesion gaps in the text (i.e., high knowledge readers). This result replicates findings with adolescent readers' (McNamara et al., 1996) and adult readers (McNamara, 2001; McNamara & Kintsch, 1996). It confirms that text understanding can be improved when the reader is induced to make more inferences and when those inferences are likely to be successful.

We further predicted that there would be an advantage for high cohesion text. While there was a main effect of cohesion for multiple-choice question performance, the benefits of cohesion were moderate and inconsistent across measures, and did not occur for science texts. To some extent, this result was predicted in the sense that we predicted an interaction of cohesion, genre, and knowledge levels. Nonetheless, the results suggest that our manipulations to the cohesion of these texts may not have been sufficient to fully support the children's understanding of the text. Prior research has clearly established a benefit of cohesion manipulations to texts for children of this age (Beck, McKeown, Omason, & Pople, 1984; Beck, McKeown, Sinatra, & Loxterman, 1991; Loxterman, Beck, & McKeown, 1994). However, the studies that have shown benefits of cohesion included manipulations to the explanatory content in the text. In contrast, in the current study, we were conservative in our approach to avoid adding large amounts of explanatory information to the high cohesion text that could not be inferred in the low cohesion text. This approach is motivated by the goal of not confounding cohesion manipulations with the addition of extra information. However, it seems that the young readers in this study needed more background information in the texts to support their understanding. Simply adding cohesion cues, and not explanatory information, is not likely to be sufficient for young readers as an approach to improving comprehension of challenging texts.

Limitations and Future Directions

The interpretation of these findings should be tempered by at least two limitations. First, although we believe that we selected the two most consistent, powerful, and theoretically valid aptitudes for reading comprehension, other reading aptitudes deemed important in previous research, such as oral language-based abilities (Carver & David, 2001; Storch & Whitehurst, 2002), metacognitive strategy knowledge (Baker, 1985; Wong, 1985), and knowledge of text structure (Williams et al., 2005), were omitted from this study. Second, although our sample was quite diverse in terms of gender, race, and socio-economic status and although they were recruited from four public schools in a large metropolitan school district, these samples do not well demonstrate representativeness of all readers in fourth grades. In fact, the children demonstrated somewhat above average performance and somewhat restricted range on the standardized, norm-based screening measures and on the standardized, norm-based reading decoding and world knowledge measures. Subsequent studies should ensure that there are sufficient readers with low normative levels on such measures.

Conclusions

The results of this study further our understanding of factors that may lead to comprehension problems for children in the fourth grade. Essentially, at this age, children are expected to be and generally are developmentally ready to make inferences while reading text and to learn from text. It is at this age that decoding skills are expected to be largely in place and it is at this age when these young children are increasingly expected to read and learn from expository texts. However, generating inferences depends on prior knowledge. If the reader does not possess sufficient knowledge, inferences are likely to be unsuccessful. Thus, comprehension problems will become more evident (in comparison to their peers), for children in the fourth grade who have not gained sufficient knowledge about the world. Decoding skills are clearly important, but deficits in decoding skills are likely to be evident on all types of texts. Thus, decoding skill problems are likely to be detected far before the fourth grade. In contrast, different problems will occur for readers who are expected to learn from text. When readers are able to make inferences, and the text calls for those inferences, differences in levels of knowledge will be a principle factor contributing to comprehension problems.

• • •

Acknowledgment
The research was supported in part by the Institute for Education Sciences (IES R305G020018-02). 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 IES. This project was a collaborative endeavor; hence we are also grateful to all of the members of the Coh-Metrix project who contributed to this study in various ways. We are particularly grateful to Rachel Best who helped to conduct and analyze this study. We are also grateful to Max Louwerse and Art Graesser who were co-investigators on this project.

Danielle S. McNamara is a Professor in the Psychology Department and Senior Scientist in the Learning Sciences Institute at Arizona State University. Her academic background includes a *Linguistics B.A. (1982), a Clinical Psychology M.S. (1989), and a Cognitive Psychology Ph.D.* (1992). Her research involves the study of learning and cognitive sciences and the development of educational technologies (Coh-Metrix, iSTART, Writing Pal; see soletlab.com). One focus of her work is on developing methods to improve success for struggling students. She has served on numerous funding panels, governing boards, editorial boards, and as Associate Editor for three journals.

Dr. Yasuhiro Ozuru is an Assistant Professor of Psychology at the University of Alaska Anchorage. He received a Ph.D. in experimental psychology with focus on cognitive psychology from New School University in 2003. He has worked on various projects related to computer-based tutoring for reading comprehension strategies and projects to develop assessment tools for elementary school students' digital literacy. His current research interests are in the areas of discourse processing and argumentation.

Randy G. Floyd is an Associate Professor of Psychology at The University of Memphis. He is the Editor of the Journal of School Psychology, and he has served on the editorial boards of Journal of Psychoeducational Assessment, Journal of School Psychology, School Psychology Review, and Psychology in the Schools. His research interests include the structure, measurement, and correlates of cognitive abilities; aptitudes predicting reading comprehension; the technical properties of early numeracy measures; and the process of professional publication.

References

Afflerbach, P. (1986). The influence of prior knowledge on expert readers' importance assignment processes. *National Reading Conference Yearbook, 35*, 30-40.

Baker, L. (1985). How do we know when we don't understand? Standards for evaluating text comprehension. In D. L. Forrest-Pressley, G. E. MacKinnon, & T. G. Waller (Eds.), *Metacognition, cognition, and human performance* (Vol. 1, pp. 155-205). Orlando, FL: Academic.

Beck, I., McKeown, M., & Gromoll, E. (1989). Learning from social studies texts. *Cognition and Instruction*, *6*, 99-158.

Beck, I. L., McKeown, M. G., Omason, R. C., & Pople, M. T. (1984). Improving the comprehensibility of stories: The effects of revisions that improve coherence. *Reading Research Quarterly, 19*, 263-277.

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.

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29,* 137-164.

Britton, B. K., & Gulgoz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology, 83*, 329-345.

Cain, K., Oakhill, J. V. & Bryant, P. E. (2000). Phonological skills and comprehension failure: A test of the phonological processing deficit hypothesis. *Reading and Writing*, *13*, 31-56.

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.

Carver , R. P., & David, A. H. (2001). Investigating reading achievement using a causal model. *Scientific Studies of Reading, 5*, 107–140.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Coté, N., & Goldman, S. R. (1999). Building representations of informational text: Evidence from children's think-aloud protocols. In H. Van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 169-193). Mahwah, NJ: Erlbaum.

Ehri, L. C. (1991). Development of the ability to read words. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. II, pp. 384-417). New York: Longman.

Graesser, A.C. & McNamara, D.S. (2010). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*.

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. Sweet and C.E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82-98). New York: Guilford Publications

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., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371-95.

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*,103-28.

Kamberelis, G., & Bovino, T.D. (1999). Cultural Artifacts as Scaffolds for Genre Development. *Reading Research Quarterly, 34, 138*–170.

Kendeou, P., van den Broek, P., White, M.J., & Lynch, J.S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*, 765-778.

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

Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. *Discourse Processes, 16,* 193–202

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

Langer, J. A. (1986). *Children reading and writing: Structures and Strategies*. Westport, CT: Ablex Publishing.

Loxterman, J. A., Beck, I. L., & McKeown (1994). The effects of thinking aloud during reading on students' comprehension of more or less coherent text. *Reading Research Quarterly, 29*, 353-367.

Lyon, G.R. (2002). Reading development, reading difficulties, and reading instruction: Educational and public health issues. *Journal of School Psychology, 40,* 3‐6.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical Manual*. *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

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., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47*, 292-330.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22*, 247-288.

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.

Meichenbaum, D., & Biemiller, A. (1998). *Nurturing independent learners: Helping students take charge of their learning*. Cambridge, MA: Brookline.

Nelson, K. (1996). *Language in cognitive development: The emergence of the mediated mind.* New York: Cambridge University Press.

Oakhill, J. V, Cain, K., & Bryant, P. E. (2003). Dissociation of single-word reading and text comprehension skills. *Language and Cognitive Processes*, *18,* 443-468.

Oakhill, J. V., Cain, K., & Yuill, N. M. (1998). Individual differences in children's comprehension skill: Towards an integrated model. In C. Hulme & R. M. Joshi (Eds.), *Reading and spelling: Development and disorder*. Mahwah, NJ: Erlbaum.

Olson, M. W. (1985). Text type and reader ability: The effects on paraphrase and text-based inference questions. *Journal of Reading Behavior, 3*, 199-214.

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes, 43,* 121-152.

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

Rupley, H. R., & Wilson, V. L. (1996). Content, domain, and world knowledge: Relationship to comprehension of narrative and expository text. *Reading and Writing, 8*, 419-432.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 6*, 934–947.

Sweet, A. P., & Snow, C. E. (Eds.). (2003). *Rethinking reading comprehension.* New York: Guilford Press.

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.

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 Press.

Vellutino, F. R. Scanlon, D. M., & Spearing, D. (1995). Semantic and phonological coding in poor and normal readers. *Journal of Experimental Child Psychology, 59,* 76 -123.

Vidal-Abarca, E., Martinez, G., & Gilabert, R. (2000). Two procedures to improve instructional text: Effects on memory and learning. *Journal of Educational Psychology*, *92,* 107–116.

Williams, J. P., Hall, K. M., Lauer, K. D., Stafford, K. B., De Sisto, L. A., & deCani, J. S. (2005). Expository text comprehension in the primary grade classroom. *Journal of Educational Psychology, 97*, 538-550.

Wong, B. Y. L. (1985). Issues in cognitive-behavioral interventions in academic skill areas. *Journal of Abnormal Child Psychology, 13*, 425-442.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III.* Itasca, IL: Riverside.

Wolfe, M. B. W. & Goldman, S. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction, 23,* 467-502.

Zinar, S. (1990). Fifth-graders' recall of propositional con- tent and causal relationships from expository prose. *Journal of Reading Behavior, 22*, 181-199.

Appendix A

Eight Texts used in Study

Effects of Heat: Low Cohesion

Moving Heat

Heat can move from one object or place to another. Heat moves from warm objects to cooler ones. You can warm your hands by holding a cup of warm soup. Heat moves from the soup through the cup to your hands. You can feel warm air rising above the cup.

Heat moves through some materials more easily than others. Heat moves easily through conductors. Most metals are good conductors. Metal pots are used for cooking. Heat from the stove quickly moves through the metal. The heat warms the food.

Other materials are not good conductors. But they may be good insulators. Insulators help keep heat from passing through. Most plastics are good insulators. So are clothes you wear, like sweaters and coats. You wear these clothes to keep warm when it is cold outside.

Changing Matter

Adding or taking away heat can change matter. Matter is something that takes up space. Matter can change from one state, or form, to another.

An ice cube is solid water. Solid is one state of matter. Heat can melt an ice cube. The ice cube changes into liquid water. Liquid is another state of matter. When heat is taken away, the water can change back. Liquid water turns into solid water.

Heat can make liquids boil. Water boils when it is heated. When the water boils, it turns into a gas. This gas is called water vapor. Solid, liquid and gas are three states of matter.

Heat from the sun causes liquid water to turn into water vapor. Water vapor mixes with the air. This is called evaporation.

Sometimes heat causes changes that cannot be changed back.

Bread can change into toast when you heat it. Eggs change when you cook them in a pan. You cannot untoast a piece of toast. You cannot uncook an egg.

Changing Air

Heat can warm air, too. A balloon is filled with air. When heat warms the air in the balloon, the air changes. The air takes up more space.

Heat from the sun warms objects all around you, like rocks, streets, and buildings. These objects then warm the air. Warm air is lighter than cold air. Warm air goes up. Cold air takes its place.

You can tell how hot or cold the air is. Temperature is a measure of how hot something is. People use thermometers to measure the temperature.

Effects of Heat on Objects, Matter, and Air: High Cohesion

Heat Moves

Heat can move from one object to another object, or it can move from one place to another place. Heat moves from warm objects to cooler ones. For example, you can warm your cold hands by holding a cup of warm soup. Your hands become warmer because heat moves from the soup, through the cup, to your hands. The heat from the soup also moves above the cup, so you can feel warm air rising above the cup.

Heat moves through some materials more easily than other materials. Conductors are materials through which heat moves easily. Most metals are good conductors. For example, metal pots are used for cooking because heat from the stove quickly moves through the metal pots and the heat in the pot warms the food.

Other materials are not good conductors, but instead are good insulators. Insulators are materials that help keep heat from passing through. For example, most plastics are good insulators. Other good insulators are the clothes you wear, especially sweaters and coats. You wear these insulating clothes in order to keep warm when it is cold outside.

Heat Changes Matter

Adding heat or taking away heat can change matter. Matter is something that takes up space. Matter can change from one state to another state, or from one form to another form. Three states of matter are solid, liquid and gas. For example, an ice cube is solid water. Heat can melt an ice cube, causing the ice cube to change into liquid water. When heat is taken away, the liquid water can change back into solid water (ice).

Heat can make liquids boil and change into a gas state. For example, water boils when it is heated. As the water boils, it turns into a gas state that is called water vapor. Heat from the sun causes liquid water to turn into water vapor. Water vapor then mixes with the air in a process called evaporation.

However, sometimes heat causes changes that cannot be changed back. As one example, bread can change into toast when you heat the bread. However, you cannot untoast a piece of toast by taking away heat. As another example, eggs change when you cook them in a pan, but of course you cannot uncook an egg by taking away the heat.

Heat Changes Air

Just as heat can warm liquids, it can also warm the air. Air is changed when it is heated. For example, if heat warms air in a balloon, the air changes by taking up more space.

Heat from the sun warms objects all around you, including rocks, streets, and buildings. These objects then warm the air. Warm air is lighter than cold air. Therefore, warm air moves upward. When the lighter, warm air goes up, the heavier, cold air moves downward. This cold air takes the place of the air that was warmed.

You can measure how hot or cold the air is by using temperature. Temperature is a measure of how hot something is. People use thermometers to measure temperature.

The Needs of Plants: Low Cohesion

What Are the Needs of Plants?

Like all living things, plants have certain needs. Plants need sunlight, water, and air to live. Plants also need minerals (MIN·uhr·uhlz). A mineral is a naturally occurring substance that is neither plant nor animal.

The parts of plants help them to get or make what they need. All plants get water and minerals from the soil. The root is the part of the plant that grows underground. Roots help hold the plant in the ground. Roots also help take in water and minerals that the plant needs.

The stem is the part that supports the plant. It helps the plant stand upright. It carries minerals and water from the roots. It also carries food from the leaves to other parts of the plant.

Some plants, such as mosses, are simple plants. They don't have real roots or stems. These plants do not grow tall. Instead, they form low-growing mats in damp places to get water directly from the soil.

Other plants, such as the redwood tree, have many roots and a large stem. They can grow very tall.

Why Does a Plant Need Leaves?

The leaves (singular, *leaf*) are the main food-making part of the plant. Many leaves have broad, flat surfaces that help them take in sunlight. Leaves are green because of *chlorophyll*. Chlorophyll traps the energy (EN·uhr·jee) in sunlight for the plant.

The leaf also helps the plant get the air it needs. Each leaf has tiny holes that take in air for the plant. The leaf uses a gas in the air called *carbon dioxide*. The plant uses the Sun's energy to combine carbon dioxide and water to make food. The stem then carries the food to the other parts of the plant.

Plants use the food they make to stay alive. When we eat plants or other animals that eat plants, we use this food, too.

When leaves make food for a plant, they give off oxygen (AHK·suh·juhn). Oxygen is a gas that is in air and water. People and animals need oxygen to live. You inhale the oxygen made by a plant with each breath you take.

How Do Plants Respond to Their Environment?

Have you ever seen a plant leaning toward a sunlit window? This is one way that plants respond to their environment. A plant responds to light by growing toward it. Some trees and shrubs drop their leaves as the days grow shorter and colder. These plants respond to the change in seasons from summer to winter.

Plants also respond to other things in the environment. For example, roots may grow toward water. The ability to respond to the environment helps a plant to live, grow, and meet its needs.

The Needs of Plants: High Cohesion

What Plants Need

Plants have certain needs, just like all living things have needs. For example, plants need sunlight, water, and air to live. Plants also need minerals (pronounced as MIN·uhr·uhlz). A mineral is not a plant or an animal. Instead, a mineral is a substance in the ground that occurs naturally. There are three parts of plants that help plants get what they need or help plants make what they need.

The Three Parts of a Plant

The three parts of the plant are the roots, stems, and leaves.

1. The Root

The root is the part of the plant that grows underground. All plants get water and minerals from the ground, which is sometimes called soil. Roots help the plant take in water and minerals that the plant needs from the soil. Roots also help hold the plant in the ground.

2. The Stem

The stem is the part that supports the plant. The stem helps the plant stand upright. It carries minerals and water from the roots of the plant to other parts of the plant. The stem also carries food from the leaves to other parts of the plant.

Some plants, such as the redwood tree, can grow very tall because they have many roots and a large stem. Other plants don't have real roots or stems. These plants are simple plants. An example of these simple plants is mosses. These simple plants do not grow tall. Instead of having roots and stems, they form low-growing mats in damp places. Simple plants get water directly from the soil through these mats.

3. The Leaves

The leaves help the plant make its food. The leaves need sunlight, air, and water to make food. Many leaves have broad, flat surfaces. These surfaces are broad and flat in order to help the leaves take in lots of sunlight. The energy in sunlight is trapped by the leaf by a substance called *chlorophyll* (pronounced KLO ro fill). Leaves are green because of chlorophyll.

The leaf also helps the plant get the air it needs to make food. This process is helped by tiny holes in each leaf. These holes take in air for the plant. The leaf only uses a gas in the air called *carbon dioxide* (CAR bun di OK side). However, the plant needs both carbon dioxide and water to make food. The plant uses the Sun's energy to combine the carbon dioxide and water to make food. The stem then carries this food to the other parts of the plant.

Plants use the food they make to stay alive. When we eat plants or when we eat animals that have eaten plants, we also use this food. That means that the food that keeps the plant alive, keeps us alive too. We also need oxygen (AHK·suh·juhn) to stay alive. The leaves give off oxygen when they make food for the plant. Oxygen is a gas that is in air and in water. All people and animals need oxygen to live. Therefore, we need plants because we inhale oxygen made by plants with each breath that we take.

How Plants Respond to Their Environment

Consider a plant leaning toward a window on which the sun shines. A plant responds to light by growing toward it. This is one way that plants respond to their environment. But plants also respond to other things in the environment. For example, roots may grow toward water. And, some trees and shrubs drop their leaves as the days grow shorter and colder in the fall. These plants are responding to the change in seasons from summer to winter. The ability of a plant to respond to the environment helps a plant to live, to grow, and to meet its needs.

Moving: Low Cohesion

John was very upset when his family moved to another state last month. He knew he would miss all his friends from the neighborhood – the twins, Charlie and Bob; Lisa, Debbie, Mike, and Jimmy. He would miss the playground two blocks from his home where they played baseball. He knew he would also miss the lake where he and his father went fishing, his friends and teachers at school, his Aunt Lucy who lived three blocks away, and the stores where he went shopping with this family. And he knew he would miss his home: the nights in the den watching sports, the barbecue parties in the backyard, his hideout in the attic and, of course, his room.

On moving day, John sat on the steps of his front porch and barely moved. "Come on," his mother said. "Help us bring some things into the car."

John could not even hear her. He was thinking of all the good times he had in the neighborhood. When some of his friends came by to say good-bye, tears flowed down his face, and he could barely look at them.

"Are you going to write to me?" he asked.

"Of course," they said. They each made sure they had the other's address and telephone number.

John and his dog, Ralph, were the last to get into the car. As his father drove away, John looked back at his house as it got smaller and smaller. When they made a left turn onto another street, he stared out the window at the familiar houses, the same ones he saw when he rode his bicycle through the neighborhood.

"Don't be upset," his mother said. "You'll meet new friends where we are going."

"Just think of it as an opportunity to meet new people," his father said. "Life is full of changes."

"And you can still keep in touch with your old friends," his mother said.

"I do not want any new friends!" John said.

John was upset for almost a week after he moved. But then he started meeting other children in his neighborhood. His neighbors had two children, Samantha and Tom, that were John's age. He met them when he played outside with Ralph. They liked a lot of the same things John liked: baseball, bicycle riding, playing tag. They introduced him to other children. In the fall, John met even more children at school. He liked his new teachers. But he still could not wait for Thanksgiving to visit his aunt and uncle in his old neighborhood, where he would see his old friends again.

Moving: High Cohesion

Last month John and his family moved to another state. John was very upset because he knew he would miss all his friends from the neighborhood. He had many friends: the twins, Charlie and Bob; Lisa, Debbie, Mike, and Jimmy. He would miss many other people too: his friends and teachers at school and his Aunt Lucy who lived three blocks away. He would miss places too. He would miss the playground two blocks from his home where he and his friends played baseball. He knew he would also miss the lake where he and his father went fishing and the stores where he went shopping with this family. And he knew he would miss his home: the nights in the den watching sports, the barbecue parties in the backyard, his hideout in the attic and, of course, his own room.

On moving day, John sat on the steps of his front porch and he was so upset that he barely moved. His mother said, "Come on, help us bring some things into the car." But, John could not even hear his mother because he was thinking of all the good times he had in the neighborhood.

Later, when some of his friends came by to say good-bye, tears flowed down John's face. Because he was so sad, he could barely look at them.

John wanted to keep in touch with his friends, so he asked, "Are you going to write to me?"

His friends said: "Of course, we will write to you." He and his friends each made sure they had each other's address and telephone number, so they could write and call.

Then it was time to leave. John and his dog, Ralph, were the last to get into the car. As his father drove away, John looked back at his house. While they were driving away it seemed that the house got smaller and smaller as it got further away. When they made a left turn onto another street, John stared out the window at the familiar houses. They were so familiar because these houses were the same ones he saw when he rode his bicycle through the neighborhood.

His parents knew he was upset and tried to comfort him. His mother said: "Don't be upset, you'll meet new friends where we are going."

His father said: "Just think of this move as an opportunity to meet new people. We know that this move is a change for you. But changes can be good. Life is full of changes."

His mother said: "You can make new friends, and you can still keep in touch with your old friends."

John disagreed, "I do not want any new friends!"

John was upset for almost a week after he moved to the other state. But then he started meeting other children in his neighborhood. His neighbors had two children, Samantha and Tom, who were John's age. He met Samantha and Tom when he played outside with his dog, Ralph. Samantha and Tom liked to do a lot of the same things John liked to do: baseball, bicycle riding, playing tag. John made more friends when Samantha and Tom introduced him to other children. In the fall, when school began, John met even more children at school. He also liked his new teachers. But he still could not wait for Thanksgiving to visit his aunt and uncle in his old neighborhood, where he would see his old friends again.

Orlando: Low Cohesion

Salvador was upset. He told Mama he was going out. He didn't want her to be worried or sad. He just needed to be alone.

He hurried across the cornfield. He turned the corner of the fort and ran through the door. No one would see or hear him cry.

The floor was just dirt, but Salvador liked his fort. He had made it himself the year before. He had brought wood scraps from the old barn behind his house.

Salvador lay down on the dirt floor. He closed his eyes. He had a picture in his mind of his mama. He thought about what happened at home before he left. He cried some more.

"We have to sell Orlando," Mama had said.

Salvador had put down his fork.

"What did you say, Mama?"

Mama tried to explain. But she could not bring herself to say more. She just looked away.

"Why would we sell Orlando?" Salvador had asked.

"You know we are moving, Salvador. And we will not have room for him in our new place. I know he is a fine pig," Mama had said.

"Orlando is more than a pig!" Salvador had shouted. "He is family."

"I know, son," Mama had said.

"If anyone tries to take poor Orlando, I won't let them!" Salvador had shouted. Then he had hurried from the house to his fort.

Salvador adored Orlando. Orlando was four years old. Salvador had named him after a place in Florida he had read about. There could not be a more perfect pet for Salvador.

As Salvador lay in the fort crying, it began to rain. At first it was only a light rain. Then it began to pour. Next there was lightning and thunder. It was a very bad storm!

Next Salvador heard a loud *crack*! Lightning had struck a tree in the forest. The tree was on fire!

Salvador got up and ran out of the fort. It was dark! He couldn't see.

He did not know which way was home.

Before long Salvador heard a snorting sound. It was Orlando! Orlando had come to find him! Salvador reached out for Orlando. He felt the pig's ear. He held on.

Orlando led Salvador home. They ran through the cornfield as fast as they could go. Mama stood by the door on the porch.

"Oh, Salvador. I was so worried!" she cried.

"Orlando saved me," Salvador said. "Yes, I know," said Mama. "He is family! We will just have to find room for him at our next place."

This brought a smile to Salvador's face. Orlando was saved!

"Now he really is a member of our family!" said Salvador.

Orlando: High Cohesion

Once upon a time, there was a boy. His name was Salvador. Salvador adored his pet pig named Orlando. Orlando was four years old. Salvador had named Orlando after a place in Florida that he had read about. There could not be a more perfect pet for Salvador than his pet pig.

One day, Orlando was eating dinner at home with his Mama. Mama said, "We have to sell Orlando."

Salvador put down his fork. "What did you say, Mama?"

Mama tried to explain, but she could not bring herself to say more. Mama just looked away.

Salvador asked, "Why would we sell Orlando?"

Mama answered, "You know we are moving, Salvador. But we will not have enough room for Orlando in our new place. We cannot take him, even though I know he is a fine pig."

Salvador shouted, "Orlando is more than a pig! He is family."

Mama said, "I know, son."

Salvador shouted, "If anyone tries to take poor Orlando, I won't let them!"

Salvador was upset, so he needed to be alone. Salvador told Mama he was going out because he didn't want her to be worried or sad.

Salvador left the house and hurried across the cornfield. He turned the corner to a fort that he had made and ran through the door. He went to the fort so no one would see him cry or hear him cry.

The floor of the fort was just dirt, but Salvador liked his fort. He had made it by himself one year ago. He had brought wood scraps from an old barn behind his house to build the fort.

After Salvador ran into his fort, he lay down on the dirt floor and closed his eyes. He had a picture in his mind of his mama. He thought about his mama telling him that they had to sell Orlando, and that made him cry some more.

As Salvador lay in the fort crying, it began to rain. At first it was only a light rain. Then it began to pour. Next there was lightning and thunder because it was a very bad storm!

Then lightning struck a tree in the forest, so Salvador heard a loud *crack*! The lightning had caught the tree on fire! Because of the noise and fire, Salvador got up and ran out of the fort. It was dark, so he couldn't see and did not know which way was home. Salvador became lost in the dark.

Before long Salvador heard a snorting sound. The snorting was from Orlando! Orlando had come to find him! When Salvador reached out for Orlando, he felt the pig's ear. Salvador held on to the ear as Orlando led Salvador home. They ran through the cornfield as fast as they could go.

When Salvador and Orlando got home, Mama stood by the door on the porch. Mama cried, "Oh, Salvador. I was so worried!"

Salvador explained, "Orlando saved me."

Mama said, "Yes, I know. Orlando must be family because he saved you! We will just have to find room for him at our next place."

So Orlando was saved, which brought a smile to Salvador's face. Salvador said, "Now Orlando really is a member of our family!"

This page is intentionally left blank

International Electronic Journal of Elementary Education, 2011, 4(1), 259-268.

Preparing for reading comprehension: Fostering text comprehension skills in preschool and early elementary school children

Paul van den BROEK*
Leiden University, The Netherlands & University of Minnesota

Panayiota KENDEOU
Neapolis University Pafos, Cyprus

Sandra LOUSBERG
HCO, The Netherlands

Gootje VISSER
Leiden University, The Netherlands

Abstract

To understand what they read or hear, children and adults must create a coherent mental representation of presented information. Recent research suggests that the ability to do so starts to develop early –well before reading age- and that early individual differences are predictive of later reading-comprehension performance. In this paper, we review this research and discuss potential applications to early intervention. We then present two exploratory studies in which we examine whether it is feasible to design interventions with early readers (3rd grade) and even toddlers (2-3 years old). The interventions employed causal questioning techniques as children listen to orally presented, age-appropriate narratives. Afterwards, comprehension was tested through question answering and recall tasks. Results indicate that such interventions are indeed feasible. Moreover, they suggest that for both toddlers and early readers questions *during* comprehension are more effective than questions *after* comprehension. Finally, for both groups higher working memory capacity was related to better comprehension.

Keywords: Reading comprehension, early intervention

Introduction

Children who experience difficulties reading and understanding information that is presented to them tend to suffer from problems in school and in their communities (e.g., NCES, 2010; NICHD, 2000). Not only are school experiences difficult for these children, but many aspects of their lives are severely hindered as they grow older. Thus, it is essential that

* Paul van den Broek, Department of Education and Child Studies Wassenaarseweg 52 2333 AK Leiden The Netherlands. E-mail: broekpwvanden@fsw.leidenuniv.nl

ISSN:1307-9298

www.iejee.com

we understand the processes that lead to successful reading comprehension and the ways in which these processes can be developed in young children. Such understanding can have far-reaching implications for educational practice, particularly with respect to assessment, diagnosis, and early intervention of reading difficulties.

In this paper, we first discuss the cognitive processes underlying reading comprehension, the development of reading comprehension, and difficulties experienced by struggling readers. We then present the results of two targeted, small-scale intervention studies aiming to support the development of reading comprehension skills in young children. Finally, we discuss the theoretical and practical implications derived from the results of this research.

Cognitive processes in reading comprehension

A common theme that has emerged from research examining the cognitive processes of reading comprehension is that, to comprehend a text, it is essential that a reader be able to decode language units and to construct a coherent mental representation of the text (e.g., Gernsbacher, 1990; Graesser, Singer, & Trabasso, 1994; Kintsch & van Dijk, 1978; Trabasso & van den Broek, 1985; van den Broek, 1994; Wagner, Piasta, & Torgesen, 2006; Whitehurst & Lonigan, 1998; Zwaan & Rapp, 2006). This representation can be accessed by the reader for different purposes after reading is completed: to recall information from the text, to answer questions, to apply the knowledge obtained from the text, and so on.

Investigations of the construction of coherent representations focus on the cognitive processes involved during reading comprehension itself, as they occur moment-by-moment when the reader proceeds through a text (*online*), as well as the resulting representations (*offline*) once reading has been completed. The online construction of a coherent mental representation of the text involves a complex set of processes such as connecting and integrating the text information that the reader currently is reading with information that occurred earlier in the text as well as with information from background knowledge. Some of these processes are quick, automatic, and relatively effortless, whereas others are slow, strategic, and relatively effortful. These processes are constrained by the limitations of working memory and by the *standards of coherence* that a reader attempts to maintain in a particular reading situation (e.g., van den Broek, in press; van den Broek, Risden, & Husebye-Hartmann, 1995; van de Velde, 1989).

The offline memory representation of the text and relevant background knowledge emerges from the processes and strategies readers employ during reading. Thus the offline representation and the online processes are causally related: The processes that unfold during moment-by-moment reading comprehension provide the basis for the construction of the offline text representation. If the online processes fail, so does the final text representation.

Development of reading comprehension

Research on reading development in preschool and early elementary-school children has provided important insights about the nature of the development of comprehension skills at this early age (for reviews of research on developmental changes in comprehension skills, see, for example, Applebee, 1978; van den Broek, 1997; van den Broek, Bauer, & Bourg, 1997; van den Broek & Kremer, 1999).

Specifically, investigations of children's memory for narratives show that children at an early age can and do engage in inferential processes, identifying meaningful relations and establishing coherence (Trabasso, Secco, & van den Broek, 1984). For example, when 4- and 6-year old children watch television programs, they tend to recall events with many causal connections better than events with fewer causal connections (van den Broek, Lorch, &

Thurlow, 1996). However, there are also systematic age differences. For instance, older children are more sensitive to the causal structure of a story than are younger children. In addition, older children are able to identify causal relations that span across multiple episodes and events in a story, whereas younger children tend to focus on recalling causal relations that occur within a particular episode of a story. Identifying causal relations across multiple episodes of a story leads to a broader and deeper understanding, for example of the story's overall theme (Williams, 1993). Furthermore, whereas older children are able to notice relations involving more abstract components such as underlying intentions and goals of characters, younger children tend to emphasize relations between concrete events in a story (van den Broek, 1997; Williams, 1993). These developmental trends reflect at least two major factors of change as children grow older: the different experiences that children have had and the increasing efficiency of their working memory.

Research on the development of language comprehension skills from preschool to early elementary school and their relation to beginning reading comprehension has shown that language comprehension and decoding skills are strongly interrelated in preschool but that their relation is weaker in kindergarten and first grade (Kendeou, van den Broek, White & Lynch, 2009). Importantly, language comprehension skills independently predict a child's reading comprehension over and above decoding skills (see also Storch & Whitehurst, 2002). The results of several other studies also highlight the importance of language comprehension skills in early reading development (Bishop & Adams, 1990; Cain & Oakhill, 2007; Cain, Oakhill, & Bryant, 2004; Catts, Fey, Zhang, & Tomblin, 1999; Paris & Paris, 2003). These findings have important theoretical and practical implications with respect to assessment in the early years of schooling, as well as interventions.We turn to this issue next.

Interventions to enhance reading comprehension

There is an impressive body of literature on reading comprehension interventions (Gersten, Fuchs, Williams, & Baker, 2001; Swanson & Hoskyn, 2001). Recent reviews of this literature have highlighted that reading comprehension improves when readers are explicitly taught various strategies such as activating prior knowledge, self-monitoring, summarizing, identifying text structures, and questioning (Faggella-Luby & Deshler, 2008). In the context of the present set of studies, we focus on questioning strategies. Questioning is an effective way to help readers construct a coherent representation of a text because it directs readers' attention to making essential connections for coherence. Existing research provides important guidance about the type of questions to ask, and about the appropriate timing of those questions.

With respect to the type of questions, it is important to consider what types of information are most useful for establishing coherence. As mentioned before, causal relations have been found to be particularly important for establishing coherence (e.g., Goldman & Varnhagen, 1986; Trabasso & van den Broek, 1985; Trabasso, van den Broek & Liu, 1988). The causal relations that readers must infer are many and complex and may extend over long distances in the text. Thus, causal questions can potentially direct the reader to attend to specific causal information in the text. Indeed, when asked causal questions, readers recall more causally connected events in narrative text than they recalled when asked general questions (McMaster, van den Broek, Espin, White, Rapp, Kendeou, Bohn-Gettler & Carlson, 2011).

With respect to the timing of questions, different questioning techniques have been used during reading to assist readers in identifying the relations that are necessary for constructing a coherent representation of the text (e.g., Hansen & Pearson, 1983; McKeown, Beck, & Blake., 2009; Trabasso & van den Broek, 1985; van den Broek et al, 2001; Yuill &

Oakhill, 1988). Consistent with the notion, mentioned above, that the cognitive processes during reading are the major determinants of success or failure in reading comprehension, it appears that for relatively proficient readers, questioning techniques *during* reading are more effective than questions *after* reading has been completed. However, interrupting the reading process with questions may be less helpful for younger children because it will direct attention and resources away from maintaining coherence (Goldman, 2004). Indeed, van den Broek et al. (2001), who investigated the timing of questioning in 4th, 7th, 10th grade and college students, showed that questioning during reading is beneficial for 7th grade and college students, but for 4th grade students, the youngest group in the study, questioning after reading was most beneficial.

Based on the findings on early development of comprehension and inference-making skills, one might surmise that it is possible to develop skills relevant to text-comprehension skills even before children can read. In the following pilot studies we explore whether it is possible to devise inference-fostering interventions modeled after the questioning methods used with older children at a very young age –for 8- to 9-year old 3rd grade children and for 2- to 3-year old toddlers. In addition, we explore whether factors such as timing of the questioning –during vs. after listening to the stories- matters and whether factors such as the child's working memory capacity and gender also influence comprehension at these ages.

Study 1- Intervention for Pre-Readers

Participants

Participants were 40 children aged 2-3 years in a day nursery in the Netherlands. The number of girls and boys was equal in the two question-timing conditions. The average age was 2.74 years (*SD* =.46).

Materials

All participants completed tests that assessed their working memory, and their comprehension and memory of three narratives. The comprehension and memory of the narratives was measured with two different traditional procedures for assessing story comprehension, namely comprehension questions and recall tasks (Pearson & Hamm, 2005). The narratives used were stories from two story books with pictures for young children ´Dikkie Dik` (Bruna, 2008) and 'Nijntje in the zoo' ('Nijntje in de dierentuin', Boeke, 2008). All three narratives were stories about animals and were age appropriate in linguistic complexity. The stories were 83, 98, and 287 words, 12, 16, 49 sentences long, respectively. The questions used to assess comprehension consisted of causal questions aimed at connecting events within each narrative. Narratives one and two had five each questions and narrative three had 10 questions, placed at natural points in the text (e.g., end of paragraph, end of page). Depending on the condition, questions were asked either during or after listening to the narrative. If the child did not respond, one neutral prompt was given. The questions were identical across conditions. Working memory capacity (WMC) was measured with the imitation sorting task (IST; Alp, 1994). The imitation sorting task measures the maximum amount of information children can simultaneously store and process.

Procedure

Participants were randomly assigned to one of two experimental conditions for each story. All children received all stories. Girls and boys were equally divided over the conditions. In each session, an experimenter read the narratives to the child and asked the causal questions either during or after the narrative (depending on condition). After each narrative was read and the questions asked, participants were asked to recall the whole story while looking at pictures from the narrative. Participants were tested individually and each session lasted approximately 30 minutes.

Results

Two separate ANCOVA`s were carried out with timing of questioning as the independent variable (during vs. after), working memory and gender as covariates, and children's recall and questions scores as the dependent variables, respectively. For the two dependent variables, recall and questions, we created difference scores by subtracting from a child's total score the mean score to account for potential differences in variance among the three narratives. Given the exploratory nature of the study and the small N, results at $p < .10$ are discussed.

The comprehension scores for the on-line questioning condition were higher than those for the off-line questioning condition, for both question-answering and recall. This difference was significant at the $p < .10$ level for recall (F (1, 39) = 3.27, $p = .07$, $\eta^2 = .08$). Among the covariates, children's WMC scores were positively related to their scores on the question-answering task F (1, 39) = 28.75, $p < .001$, $\eta^2 = .34$, as well as on the recall task F (1, 39) = 10.75, $p < .001$, $\eta^2 = .22$.

Study 2- Intervention for Young Readers

Participants

Participants were 42 3rd grade children aged 8-9 years old in an elementary school in the Netherlands. The number of girls and boys was equal in the two question-timing conditions. The average age was 8.61 (*SD* =.47).

Materials

All participants completed tests that assessed their working memory, and their comprehension and memory of two narratives. As in Study 1, the comprehension and memory of the narratives was measured using comprehension questions and recall tasks (Pearson & Hamm, 2005). The narratives were stories from the children's book 'Hare and Rabbit' ('Haas en Konijn', Delft, 2000). Both narratives were age appropriate. The stories were 834 and 796 words, 106 and 113 sentences long, respectively. The questions consisted of causal questions aimed at connecting events within each narrative. Each narrative had 10 questions, placed at natural points in the text (e.g., end of paragraph, end of page). The questions were identical across conditions. Working memory capacity (WMC) was measured using the using a Dutch translation of the Sentence Span Measure (Swanson, 1995).

Procedure

Participants were randomly assigned to one of two experimental conditions. Stories and conditions were counterbalanced so that across the children all stories occurred in each condition. Girls and boys were equally divided over the conditions. In each session, an experimenter read the narratives to the child and asked the causal questions either during or after the narrative (depending on condition). If the child did not respond, one neutral prompt was given. After each narrative was read and the questions asked, participants were asked to recall the whole story while looking at pictures from the narrative. Participants were tested individually and each session lasted approximately 30 minutes.

Results

Two separate ANCOVA`s were carried out with timing of questioning as the independent variable (during vs. after), working memory and gender as covariates, and children's recall and questions scores as the dependent variables, respectively. For the two dependent variables, recall and questions, we again created difference scores by subtracting from a child's total score the mean score to account for potential differences in variance among the three narratives.

Given the exploratory nature of the study and the small N, results at $p < .10$ are discussed. The comprehension scores for the on-line questioning condition were higher than those for the off-line questioning condition, for both question-answering and recall. This difference was significant at the p<.10 level for questioning-answering (F (1, 20) = 3.46, p =.07, $\eta^2 = .28$. Among the covariates, children's WMC scores were positively related to their scores on the question-answering task F (1, 39) = 14.84, $p < .001$, $\eta^2 = .52$, as well as on the recall task, F (1, 39) = 22.25, $p < .001$, $\eta^2 = .60$.

General Discussion

To understand what they read, children and adults must create a coherent mental representation of the presented information. To do so, they need to identify relations between various parts of the information, often through inferential processes. As detailed in the literature review in this paper, results from recent cross-sectional and longitudinal studies indicate that the ability to connect pieces of information and to create coherence starts to develop early –well before reading age. Moreover, they show that individual differences in comprehension skills at an early age (as young as 4 years) are predictive of later reading-comprehension performance (in the middle and upper elementary school grades), independently from decoding and other basic skills.

These findings raise the possibility that early interventions could be developed and used to put very young children on the right track towards becoming good readers. This possibility is important to explore as report after report notes that many elementary school children do not develop adequate reading comprehension skills (National Center for Education Statistics, 2010; Coulombe, Tremblay, Marchand, 2004; Verhoeven, Biemond, Gijsel, & Netten, 2007).

The findings from two exploratory studies demonstrate that it is possible to design and implement interventions aimed at fostering text comprehension skills in early readers (3rd grade) and even toddlers (2-3 years). These interventions were modeled after questioning techniques that have been employed successfully in a reading context with children in later elementary school and high school (e.g., McMaster et al., 2011). In a listening context, causal questions are used to encourage children to identify semantic relations between various parts of age-appropriate stories that are being read to them (e.g., Kendeou, Lynch, van den Broek, Espin, White & Kremer, 2005; van den Broek, Kendeou, & White, 2009).

Besides showing the feasibility of interventions at early ages, the results suggest that for both age groups questions *during* comprehension are more effective. This is consistent with the view (Rapp, van den Broek, McMaster, Kendeou & Espin, 2007) that the *process* of comprehension is where the foundation for success and failure in comprehension is determined. It is at odds with earlier findings in reading contexts (e.g., van den Broek, Tzeng, Risden, Trabasso & Basche, 2001) that 4th grade students sometimes perform better following a questioning *after* reading than following an intervention *during* reading. In these earlier studies it was hypothesized that the decoding of the text itself required a considerable portion of a child's attentional and processing capacities and that the added burden of having to read and answer questions during reading was detrimental to comprehension and memory. The current findings are consistent with this explanation: when the demands on decoding were reduced by using a listening task, the benefit of eliciting online processes re-emerged. From a practical point-of-view this suggests that comprehension-building interventions may be most beneficial when presented in a format that does not require extensive other skills such as decoding.

In studies with adult readers, an important factor affecting the ability to make inferences and comprehend a text consists of the reader's working memory capacity (Kaakinen, et al., 2003; Linderholm & van den Broek, 2002). Differences in working memory capacity also have been implicated as an important factor in determining comprehension performance by elementary school readers (Cain, Oakhill, & Bryant, 2004). The results in the present studies extend the role of working memory as a powerful factor in comprehension to an even younger age. For both 3rd grade students and toddlers differences in working memory capacity were strongly related to differences in comprehension performance, as measured by recall and questioning tasks (effect sizes of .22 and .34 for toddlers and of .55 and .60 for the 3rd grade children).

In conclusion, the findings on the effects of questioning on comprehension in 3rd grade children and toddlers show that the approach of developing early interventions using non-reading contexts aiming to enhance children's inferential and comprehension skills may have considerable promise. It would be worthwhile to engage in full-scale investigations of the effectiveness of such interventions and of properties that determine their impact, such as timing, intensity, duration, types of questions, and so on. Longitudinal studies are needed to determine long-term effects and whether they indeed lead to improved reading comprehension in later years. In the past, we have argued that it is essential to select or design interventions that address the underlying causes of the difficulties struggling readers face (Kendeou, van den Broek, White, & Lynch, 2007; Rapp et al., 2007; van den Broek et al., 2005; van den Broek, White, Kendeou & Carlson, 2009). Effective interventions are those that influence readers' actual processes during comprehension, particularly at points where children's comprehension processes tend to break down. In doing so, they also affect the memory representation and understanding that children have after they listened to a text. The current studies show that questioning techniques can be adapted to very young children and can indeed affect their comprehension. These techniques therefore provide a promising starting point for the development of interventions aimed at fostering comprehension and inference-generation skills in children well before they start to read texts.

• • •

Paul van den Broek, Ph.D., is a Professor of Education and Child Studies at Leiden University (Netherlands) and Cognitive Sciences at the University of Minnesotat. He is also the Director of the Brain and Education lab at the University of Leiden. His expertise is on cognitive processes (and their development) in reading comprehension and on reading comprehension interventions.

Panayiota Kendeou, Ph.D., is an Assistant Professor of Educational Psychology at Neapolis University Pafos, Cyprus. Her current research focuses on the cognitive processes that support memory and learning in the context of reading comprehension. She is currently an Associate Editor for the Journal of Research in Reading.

Sandra Lousberg has a master's degree in Educational Sciences from Leiden University. She works as an educational specialist at the HCO school guidance center in The Hague, the Netherlands.

Gootje Visser has a master's degree in Educational Sciences from Leiden University. She works as an educational specialist at Van de Graaf in Amstelveen, the Netherlands, an organization that specializes in e-learning.

References

Alp, I.E. (1994). Measuring the size of working memory in very young children: The Imitation Sorting task. *International Journal of behavioural Development, 17,* 125-141.

Applebee, A. (1978). Teaching high-achievement students: A survey of the winners of the 1977 NCTE Achievement Awards in writing. *Research in the Teaching of English, 1*,

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 31*, 1027-1050.

Boeke, J. (2008). *Dikkie Dik; small creatures,* Dutchbook BV, Haarlem.

Bruna, D. (2008). *Nijntje in de dierent (Nijntje in the zoo).* Mercis Publishing BV, Amsterdam.

Cain, K. & Oakhill, J. (2007). Reading comprehension difficulties.: Correlates, causes, and consequences. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems: Oral and written language* (pp. 41-75). New York: Guildford.

Cain, K., Oakhill, J. V., & 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.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3,* 331–361.

Coulombe, S., Tremblay, J-F, Marchand, S (2004). *Literacy scores, human capital and growth across fourteen OECD countries*. Ottawa: Statistics Canada.

Delft, G. (2000). *Haas en Konijn; een bos vol vrienden (Hare and Rabbit; a forest full of friends)*. Uitg. Holland, Haarlem.

Faggella-Luby, M. N., & Deshler, D. D. (2008). Reading comprehension in adolescents with LD: What we know; what we need to learn. *Learning Disabilities Research & Practice, 23*, 70-78.

Gernsbacher, M.A. (1990). Language comprehension as structure building, Hillsdale, N.J.: Erlbaum.

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*(2), 279-320.

Goldman, S. R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. Shuart-Faris & D. Bloome (Eds.), *Uses of intertextuality in classroom and educational research* (pp. 317–351). Greenwich, CT: Information Age.

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

Hansen, J., & Pearson, P. D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. *Journal of Educational Psychology, 75*, 821-829.

Kaakinen, J. K., Hyöna, J., & Keenan, J. M. (2003). How prior knowledge, WMC, and relevance of information affect eye fixations in expository text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 447-457.

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, Online First.* Retrieved from http://dx.doi.org/10.1007/s10643-005-0030-6

Kendeou, P., van den Broek, P., White, M.J., & Lynch, J. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and code-related skills. *Journal of Educational Psychology, 4*, 765–778.

Kendeou, P., van den Broek, P., White, M., & Lynch, J. (2007). Comprehension in preschool and early elementary children: Skill development and strategy interventions. In D.S. McNamara (Ed.) *Reading comprehension strategies: Theories, interventions, and technologies,* (pp.27-45). Mahwah, NJ: Erlbaum.

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

Linderholm, T. & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology, 94*, 778-784.

McKeown, M. G., Beck, I. L., & Blake, R. G. K. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly, 44*, 218-253.

McMaster; K., van den Broek, P., Espin, E., White, M.J., Rapp, D., Kendeou, P., Bohn-Gettler, C., & Carlson, S. (2011). *Making the Right Connections: Differential Effects of Reading Intervention for Subgroups of Comprehenders*.

National Center for Education Statistics (2010). The nation's report card: Reading 2009. Retrieved August 2, 2010 from http://nces.ed.gov/nationsreportcard/reading/.

National Institute of Child Health and Human Development. (2000). *Report of the national reading panel.* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Paris, A., & Paris, S. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, 38 (1), 36–76.

Pearson, D. P., & Hamm, D. N. (2005). *The assessment of reading comprehension: A review of practices past, present, and future.* In S. G. Paris & S. A. Stahl (Eds.), Children's reading: Comprehension and assessment (pp. 13–70). Mahwah, NJ: Erlbaum.

Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading, 11*, 289-312.

Storch, S.A., & Whitehurst, G.J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*, 934–947.

Swanson, H. L. (1995). *The Swanson Cognitive Processing Test*. Austin, Texas:PRO-ED.

Swanson, H. L.,& Hoskyn, M. (2001). Instructing adolescents with learning disabilities: A component and composite analysis. *Learning Disabilities Research & Practice, 16,* 109–119.

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: Erlbaum.

Trabasso, T., van den Broek, P. W., & Liu, L. (1988). A model for generating questions that assess and promote comprehension. *Questioning Exchange, 2*, 25-38.

Trabasso, T. & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24,* 612-630.

van de Velde, R.G. (1989). Man, verbal text, inferencing and coherence. In W. Heydricch, F. Neubauer, J.S. Petofi & E. Sozer (Eds.). *Connexity and coherence: Analysis of text and discourse.* New York: Walter de Gruyter.

van den Broek, P. (in press). Individual and developmental differences in reading comprehension: Assessing cognitive processes and outcomes. In J. Sabatini and E. Albro (Eds.), *Assessing Reading in the 21st Century: Aligning and Applying Advances in the Reading and Measurement Sciences*. New York: Taylor & Francis.

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 cement of the universe: The development of event comprehension from childhood to adulthood. In P. van den Broek, P. J. Bauer, & T. Bourg (Eds.), *Developmental spans in event comprehension and representation* (pp. 321-342). Mahwah, NJ: Erlbaum.

van den Broek, P., Bauer, P. J, & Bourg, T. (1997). *Developmental spans in event comprehension and representation.* Hillsdale, NJ: Erlbaum.

van den Broek, P., Kendeou, P., & White, M.J. (2009). Cognitive processes during reading: Implications for the use of multimedia to foster reading comprehension. In A.G. Bus & S.B. Neuman (Eds.), *Multimedia and literacy development: Improving achievement for young learners* (pp. 57-73). New York: Taylor & Francis.

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

van den Broek, P.W., Lorch, E.P., & Thurlow, R. (1996). Children's and adults' memory for television stories: The role of causal factors, story-grammar categories and hierarchical level. *Child Development, 67*, 3010-3029.

van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist approaches in accounts of reading comprehension. *Discourse Processes, 39,* 299-316.

van den Broek, P., Risden, K., & Husebye-Hartmann, E. (1995). The role of readers' standards for coherence in the generation of inference during reading. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353-373). Hillsdale, NJ: Erlbaum.

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.

van den Broek, P., White, M. J., Kendeou, P., & Carlson, S. (2009). Reading between the lines: Developmental and individual differences in cognitive processes in reading comprehension. In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence. (Eds.), *Beyond decoding: The behavioral and biological foundations of reading comprehension* (pp. 107-123). New York, NY: The Guildford Press.

Verhoeven, L., Biemond, H., Gijsel, M., & Netten, A. (2007). Taalvaardigheid

Nederlands: Stand van zaken in 2007. (Reading proficiency in the Netherlands: The status in 2007). Nijmegen: Expertisecentrum Nederlands.

Wagner, R. K., Piasta, S. B., & Torgesen, J. K. (2006). Learning to Read. In M. Traxler, M. A. Gernsbacher (Eds.). *Handbook of Psycholinguistics* (2nd Edition). Academic Press.

Williams, J.P. (1993). Comprehension of students with and without learning disabilities: Identification of narrative themes and idiosyncratic text representations. *Journal of Educational Psychology, 85,* 631-641.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69,* 848-872.

Yuill, N., & Oakhill, J. (1988). Effects of inference awareness training on poor reading comprehension. *Applied Cognitive Psychology, 2*, 33-45.

Zwaan, R. A., & Rapp, D. N. (2006). Discourse comprehension. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 725–764). San Diego, CA: Elsevier.