

METALINGUISTIC INSTRUCTION IMPROVES THIRD GRADERS' READING
COMPREHENSION

by

MARCY ZIPKE

A dissertation submitted to the Graduate Faculty in Educational Psychology in partial
fulfillment of the requirements for the degree of Doctor of Philosophy, The City
University of New York

2007

UMI Number: 3245042

Copyright 2007 by
Zipke, Marcy

All rights reserved.

UMI[®]

UMI Microform 3245042

Copyright 2007 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© 2007

MARCY ZIPKE

All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Educational Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Linnea Ehri, Ph.D., Chair of Examining Committee

Date

Alan Gross, Ph.D., Executive Officer

Date

David Rindskopf, Ph.D.

Helen Cairns, Ph.D.

Helen Johnson, Ph.D.

Julie Rosenthal, Ph.D.

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

METALINGUISTIC INSTRUCTION IMPROVES THIRD GRADERS' READING
COMPREHENSION

by

Marcy Zipke

Adviser: Distinguished Professor Linnea Ehri, Ph.D.

The ability to understand complex language, including syntax and semantics, is acknowledged to be an important component of reading. However, metalinguistic awareness training has not traditionally been a component of reading comprehension instruction. The purpose of this study was to investigate the efficacy of instruction in metalinguistic awareness for improving reading comprehension. An active learning program utilizing riddles and other ambiguous texts was taught to twenty-three third-graders. They learned that words, phrases, and sentences can sometimes have more than one meaning, and that meaning is dependent on context. In addition, they were taught how to determine the meaning intended in the text. A control group, made up of twenty-three participants matched for comprehension ability, received reading comprehension lessons without a metalinguistic component. The objectives were to determine (1) whether or not metalinguistic awareness is trainable in a simple and fun school-based program, and (2) whether training in metalinguistic awareness transfers directly to reading comprehension. It was found that a short period of training was effective in teaching the treatment group how to identify and define homonyms and

ambiguous sentences. Furthermore, trained students improved significantly more than the control group in their reading comprehension as measured by a standardized cloze task, as well as in their comprehension monitoring. The implications for the use of language-based methods of reading comprehension in the classroom are discussed.

Acknowledgments

First and foremost I have to thank my husband, Christopher Hasler, for believing in me when I wasn't sure I could do this. He never doubted me, and his confidence is contagious. Among his many good qualities, I also have to especially thank him for his ability to make me laugh in the face of stress. Without that, I would have collapsed years ago.

The years I spent at the Graduate Center were certainly my most rewarding as a student. I have Linnea Ehri to thank for that: her amazing depth of knowledge is an inspiration to all of her students. She is wise, patient and persistent; she pushes all of her students to go above and beyond what they think they can do. Thanks too, to all of the professors at the GC for their open-door policy and habit of funneling articles to students at the mere mention of an interest.

Special thanks to Helen Cairns, who was encouraging at all of the right moments, and who dedicated an extraordinary amount of time to tutoring me in linguistics. Also to Julie Rosenthal, Helen Johnson, and David Rindskopf for reading the dissertation and answering all of my panicked emails.

Many creative ideas that ultimately made it into this study's training tasks came from brainstorming sessions with my dissertation group. Thanks especially to Kathy Kaufman, Alba Langenthal, Mark Lauterbach, Simone Nunes, Connie Petropoulos, Alan Tener, Alicia Senia, and Julie Rosenthal (again) for all of your brilliant suggestions.

And finally, thanks to my parents, Allen and Jean Zipke, for the steady supply of encouragement, support, editing, subjects, advice, money, respite, you-name-it...

Table of Contents

Chapter 1: Introduction	1
Chapter 2: Review of the literature	8
Metalinguistic Awareness	8
Understanding Ambiguity	17
Comprehension Monitoring	28
Chapter 3: Preliminary Studies	43
Chapter 4: Dissertation Study Rationale and Design	50
Chapter 5: Method	53
Participants	53
Materials and Procedures	53
Procedure	62
Training: Metalinguistic Awareness of Multiple Meanings	62
Control Group Training: Book Reading and Discussion	67
Design and Statistical Analyses	67
Chapter 6: Results	69
Characteristics of Participants	69
Effectiveness of Metalinguistic Training: Comparison of Experimental and Control Groups	70
Response to Metalinguistic Awareness Training	75
Regression Analyses	84

Qualitative Analysis of Good, Average, and Poor Readers' Response to Metalinguistic Awareness Training by Reading Ability	89
Chapter 7: Discussion	94
Overview	94
Effects of Training in Metalinguistic Skills	95
Comprehension Monitoring Effects	97
Transfer to Reading Comprehension	99
Limitations and Future Research	104
Conclusions and Implications	106
Appendix A: Written Consent Form	108
Appendix B: Script for Obtaining Oral Consent from Participants	110
Appendix C: Homonym definition test: Metalinguistic awareness	111
Appendix D: Ambiguous sentence comprehension task: Metalinguistic awareness	112
Appendix E: Riddle resolution: Metalinguistic awareness	114
Appendix F: Heteronym pronunciation: Comprehension monitoring	116
Appendix G: Miscue self-correction: Comprehension monitoring, task passages	118
Appendix H: Error detection: Comprehension monitoring	126
Appendix I: Experimental Group Script and Materials	131
Appendix J: Control Group Script and Materials	143
References	145

List of Tables

Table	Description	Page
1	Yuill's (1996) Riddle Classification	21
2	Characteristics and Mean Performance of the Metalinguistic Treatment and Control Groups on the Screening Tests and Pretests	54
3	Mean Performance of Metalinguistic Treatment Group and Control Group on Posttests and Test Statistics	71
4	Proportion of Participants in Each Treatment Group Who Responded Correctly to Each Item on the Posttest Homonym Definition Test	76
5	Proportion of Participants in Each Treatment Group Who Responded Correctly to Each Item on the Posttest Ambiguous Sentence Definition Test	77
6	Proportion of Participants in Each Treatment Group Who Responded Correctly to Each Item on the Posttest Anomaly Detection Test	78
7	Intercorrelations Between Participants' Success with Metalinguistic Training and Verbal and Literacy Abilities at Pretest	85

METALINGUISTIC INSTRUCTION IMPROVES THIRD GRADERS' READING COMPREHENSION

Chapter 1: Introduction

Good reading comprehension is the most important capability enabling students to learn in school. Every subject requires processing texts, elaborating on texts, and remembering the information. Cornoldi and Oakhill (1996) believe that children with reading comprehension problems are slow to learn because they cannot comprehend textbooks and lectures with the same proficiency as their classmates. Nevertheless, the systematic study of reading comprehension processes and research into remediation for reading comprehension difficulties is relatively recent. The National Reading Panel (National Institute of Child Health and Human Development, 2000) reports that research into reading comprehension did not begin until the 1970s.

Markman (1977, 1979, 1981) was one of the first researchers to focus on comprehension. She studied the reader's online awareness of the comprehension process. Her initial finding was that both young and mature readers fail to monitor their comprehension. Her participants failed to detect logical and semantic inconsistencies in text. This discovery of comprehension failure led other researchers to focus on the processes that comprise text comprehension and the most effective methods for teaching comprehension.

In the past thirty years, researchers have found evidence for the impact of many other capabilities in addition to comprehension monitoring. For example, working memory, inferential ability, vocabulary, background knowledge, syntactic competence, knowledge of the reading process, and strategy usage have all been related to reading

comprehension success (Cornoldi & Oakhill, 1996). The interactive model of reading presented by Ehri (1998) and adapted from Rumelhart (1977), presents all of the above as components of successful reading. Knowledge of language, knowledge of the world, and metacognitive knowledge are postulated to be top-down processes, whereas knowledge of the graphophonemic system, a lexicon of sight words, and spoken vocabulary contribute from the bottom up. Ryan and Ledger (1984) point out that in order to become a fluent reader, the student must become adept at making meaning by applying knowledge about technical language features. In this way, children with advanced metalinguistic knowledge about sentence structure can be expected to make faster progress. Nevertheless, while knowledge of language, such as syntax and semantics, is acknowledged in the model of reading described above, training in reading comprehension has not traditionally been approached in this manner.

Although inarguably important, text comprehension has traditionally been taught in the content areas, not as part of formal reading instruction (Durkin, 1979). But Markman (1977, 1981) and others' work contributed to the idea that reading is an active procedure—intentional thinking through which meaning is constructed. Today, reading comprehension is viewed as an active process, in which the reader constructs meaning by interacting with the text. Reading comprehension constitutes “construction of the meaning of a written text through a reciprocal interchange of ideas between the reader and the message in a particular text” (National Institute of Child Health and Human Development, 2000, p.4-39). The reader constructs mental representations of what is read and stores them in memory, along with semantic interpretations that were made while reading (Kintsch & van Dijk, 1978). The formation, storing, and processing of

these representations and interpretations demand the use of many complicated processes. Failure at any step of the way may result in a failure to comprehend.

A trend in the reading comprehension research of the past thirty years has therefore been to focus on the training of particular strategies that can be consciously implemented by the reader (National Institute of Child Health and Human Development, 2000). While many researchers have had success with strategy training (e.g. Palinscar & Brown, 1984; Pressley, 2000), others have instead examined the processes that contribute to comprehension and have attempted to determine how these processes can best be fostered in beginning readers.

Reading is a language-based skill (Shankweiler, Crane & Macaruso, 1992), so problems in oral language constitute a risk factor for reading disability (Catts, 1993). Based on these facts, Roth and her colleagues (1996) have argued that more research is needed to examine the role of metalinguistic awareness in reading. Roth and her colleagues (1996, p.258) define metalinguistic awareness as “the ability to objectify language and dissect it as an arbitrary linguistic code independent of meaning.” They include, under this umbrella term, phonemic awareness, metasyntax, metamorphology, and metasemantics. They argue that the success of phonemic awareness in predicting decoding skill (e.g. Bryant & Bradley, 1985) raises the question of how other metalinguistic skills might contribute to reading. Ryan and Ledger (1984) describe metalinguistic development as a shift of attention from meaning to structure in tasks that require deliberate control over language. They identify the ability to decenter and shift one’s focus from meaning and context to structure as crucial to comprehension.

Researchers such as Tunmer, Nesdale, and Wright (1987) and Gaux and Gombert (1999) have confirmed a relationship between syntactic awareness and reading. Tunmer, Herriman, and Nesdale (1988) proposed a model of reading in which phonological awareness is most important in word recognition and syntactic awareness most important after the words have been identified and meaning is being constructed. They reason that syntactic awareness influences reading by enabling readers to monitor their comprehension (see Bowey, 1986) and by helping with word recognition by facilitating the use of context to determine unknown words.

Ambiguity instruction combines training in syntactic and semantic awareness with an awakening of the need for comprehension monitoring. Studies by Cairns, Waltzman, and Schlisselberg (2004) and Yuill (1998) have demonstrated the importance of understanding that words and sentences can have more than one meaning, as well as how to determine the appropriate meaning from context. Ambiguity training means teaching developing readers how to rearrange words within a sentence, or to stress different words within a sentence in order to apprehend additional meanings in the text. Ambiguity training enhances syntactic awareness, which in turn facilitates vocabulary learning through improved use of contextual cues. When children come to a word they cannot read in a sentence, they are flexible enough to use both their knowledge of the meaning of the sentence as well as the grammatical role of the word to figure it out (Tunmer & Bowey, 1984). This contextual facilitation, in turn, contributes to the readers' vocabulary acquisition and may result in an overall improvement in reading skill, by enabling the reader to allocate more attention to comprehension and less attention to resolving the identity of unfamiliar words in the text. Improved syntactic

processing may also enable the parsing of sentences, which in turn leads to more effective processing of ambiguities and resolution of breakdowns in comprehension, which in turn leads to speedier text comprehension (Carlisle, 1995; Leong, 1984).

As mentioned above, another related process that may benefit from training in recognizing and re-examining the meaning of ambiguous sentences is comprehension monitoring. In order to evaluate and regulate one's comprehension of text, it is necessary to know that the words in a text can add up to more than one possible meaning, and that context and alternative explanations need to be considered. Yuill (1996) used riddles to measure and teach metalinguistic tasks to 7-11-year-olds. She reasoned that comprehension monitoring is a metacognitive task, whereas riddle appreciation is specifically metalinguistic (p.195). Metacognitive and metalinguistic tasks go hand in hand, in that linguistic knowledge needs to be monitored and applied accordingly. This is what makes metalinguistic awareness and comprehension monitoring skills compatible in the form of ambiguity training.

Ryan and Ledger (1979) experimented with the idea that young children may already have a latent understanding of sentence structure, since they react to most language forms passively, rather than analytically. If this were true, then children's metalinguistic performance could be enhanced simply with a "stop and think" strategy. They hypothesized that children who do things like pause, whisper the target sentence aloud, attempt a correction, and make judgments based on the success/failure of that correction, would perform better on sentence judgment and correction tasks. Twenty-eight kindergarteners scoring below the 67th percentile on a sentence judgment and correction task participated first in a sentence judgment task that required participants to

simply deem sentences right or wrong. Next, the correction task required rearrangement (“boy the” to “the boy”) or re-composition. Six types of syntactic deviations were included: *wh*-questions, negatives, articles, irregular verbs, and separable verbs. A training group received six brief sessions, in which they were encouraged to “stop and think” in order to improve their judgments. Stimuli were graded in difficulty and included article-noun pairs (“boy the” versus “the boy”), two-word sentences, three-word phrases, three- and four-word sentences, and six-to-seven word sentences containing subtle syntactic errors. A no-treatment control group was pre- and posttested. Although the experimental group did perform slightly better on the posttest, the results were statistically insignificant. The authors interpret this as support for other findings that have shown that accuracy feedback alone does not enhance sentence judgments (i.e. Bohannon, 1975; Scholl & Ryan, 1980). Further, the authors suggest that kindergartners are not yet capable of focusing on the structure of sentences. This sort of training may be more successful with older children because they have the metalinguistic capability for the task; when older children fail, it may be due to an inappropriate task orientation rather than a lack of metalinguistic knowledge about syntax. The training study described herein therefore targets older children, and combines strategy instruction with a more meaning-oriented focus.

In the current study, although students are not taught to “stop and think,” they are encouraged to reprocess ambiguous sentences and riddles until the second meaning becomes clear. One aim is to provide evidence justifying teachers’ use of riddles and other devices employing multi-meaning language. The hypothesis tested is that students

who are taught to understand ambiguity will become more syntactically aware, will increase their comprehension monitoring, and will improve their text comprehension.

This research holds important implications for teaching. Riddles offer especially engaging instructional content for teaching as well as for assessment. Most children are familiar and comfortable with riddles. They have heard a riddle before, whether or not they understood the ambiguity in it. This contrasts with unfamiliar tasks, such as asking the child to pronounce a word without a certain sound (e.g., say “smile” without “mmm”), that are found in standardized language assessments. In addition, riddles are especially suited for young readers because they are whole texts. As short as they are, they are discrete units with meaning (Kazemek, 1999). This makes their use as a learning or assessment tool more developmentally appropriate than sentence fragments or words without a specific frame of reference. Lastly, riddles are enjoyable for all involved. Present findings show that teaching children to manipulate language, write riddles, and read ambiguous text increases their understanding of ambiguity, their comprehension monitoring, and their reading comprehension, hence suggesting the value of teaching multiple-meaning words and sentences in the classroom.

Chapter 2: Review of the literature

Metalinguistic Awareness

Three intersecting areas of research have bearing on the research at hand. The first to be discussed here is metalinguistic awareness, or the ability to play with language. Related to metalinguistic awareness is understanding ambiguity, or the process of learning to see that words and sentences have more than one meaning and how to recognize the intended meaning. Lastly, the research on comprehension monitoring will also be reviewed.

Linguists, psychologists, and reading researchers have all investigated the development of metalinguistic awareness in children's cognitive development. Numerous studies have shown that metalinguistic awareness contributes to reading comprehension, although competing theories exist to explain the nature of this contribution.

For example, Tunmer and Bowey (1984) posited that a child's initial focus in language acquisition is on what they say. This focus on meaning gets set aside when the child first learns to read and needs to focus on form in preference to meaning. It is not until decoding skills are securely in place that the child begins paying attention to the sentential context of words, rather than words in isolation. Tunmer and Bowey (1984, p.162) liken this to "putting humpty-dumpty back together again." For some children, the ability to refocus attention on meaning does not come easily. These children have not developed the metalinguistic awareness that they need to tie together the seemingly disparate skills learned in a short time. Tunmer and Bowey (1984) suggested that metalinguistic awareness—specifically syntactic awareness—facilitates the process of

text integration, leads to improved comprehension monitoring, and thereby increases comprehension.

Downing (1978) has proposed that children approach reading without being able to clearly conceptualize that the purpose of reading is the extraction of meaning, and that decoding and word recognition are merely means to the goal. Downing's idea is that metalinguistic knowledge is necessary for obtaining meaning, rather than a prerequisite for decoding. Tunmer and Bowey (1984) tempered this stance by proposing that different metalinguistic skills are distinctive in terms of decoding and comprehension. They divided different aspects of linguistic awareness into lower- and higher-level features, according to the size of the linguistic unit: According to this model, phonological skills are lower-level and influence decoding; structural or syntactic skills are higher-level and influence comprehension.

Congruent evidence for this assertion comes from Bowey (1986) who administered reading comprehension, oral reading, and syntactic awareness assessments to 48 fourth- and fifth-graders. Participants were divided into skilled and less skilled reading groups based on performance on a standardized word reading task. Syntactic awareness was measured by examining performance on error imitation and error correction tasks: the children were read sentences that contained a "mistake" so that they sounded wrong. For example, one of the practice items was, "John running is." In the imitation task, the participants were required to repeat the sentence, with the mistake. In the correction task, they were required to fix the error. The number of answers that conformed to task requirements was scored as the number correct. Reading comprehension was measured with standardized tests. In the oral reading task, the

children were given reading material and asked to read it as well as they could. The experimenter did not provide assistance. A three-way ANOVA (grade by group by task) revealed that good readers performed better than less skilled readers on the syntactic awareness tasks. Grade was not a factor. First order correlations between the syntactic measures and reading comprehension scores were significant, even when vocabulary and age effects were statistically controlled. Bowey (1986) interpreted these results to mean that the difference between skilled and less skilled readers represents a substantial delay in the development of syntactic awareness. She reasoned that syntactic awareness plays a direct role in reading comprehension through efficient comprehension-monitoring skills. Less skilled readers in this study were less able to monitor the grammatical well-formedness of the sentences and thus had greater difficulty locating the sources of comprehension difficulty in texts. She drew the following conclusion: “The present finding that less skilled decoders’ difficulties with ongoing comprehension monitoring are not restricted to written language tasks, but are also evident in oral language tasks, and that syntactic awareness is correlated with ongoing reading comprehension monitoring suggest that teaching children to monitor semantic and syntactic consistency within oral tasks may directly facilitate reading comprehension” (p.296).

Contrary to this evidence, Glass and Perna (2001) administered four tasks to fourth-graders who differed in reading skill as determined by their classroom teachers. Students were given a test of syntactic competence and three subtests of the green level of the Stanford Diagnostic Reading Test (Karlsen, Madden & Gardner, 1976): the vocabulary subtest and two forms of the reading comprehension subtest. Form A of the reading comprehension subtest was given in its printed form, following the instructions in

the manual. Form B was presented aurally, to be a measure of auditory comprehension. While the poor readers performed significantly poorer on the syntactic test than did the good readers, the difference was small. And when vocabulary was controlled, there was no correlation between syntactic competence and either auditory or reading comprehension for good or poor readers. However, only 27 poor readers participated in this study (as compared to 95 good readers), so this failure to find a correlation may be due to a lack of power. In addition, Glass and Perna used a complicated and innovative task to assess syntactic competence, in which participants had to match sentences of different degrees of syntactic complexity to one of four picture alternatives. Examples of sentences are the following: "The girl asked the policeman who pointed to it." "The policeman asked by the girl pointed to it." "The policeman the girl asked pointed to it." Performance of children who received a sentence at one level of complexity was compared with the performance of children who received another. In addition, all participants received two sentences with parallel syntactic structures so that each child's comprehension on a variety of forms could be assessed. Although this was an auditory task, the authors argue that memory load was not a factor, as performance was unaffected by the length of the sentence. But since the goal was to establish the relationship of syntactic awareness to reading disability, it is arguable that the sentences should have been presented in writing.

While Glass and Perna (2001) found no evidence for the syntactic hypothesis, Abrahamsen and Shelton (1989) approached the problem differently. They used a text manipulation approach to demonstrate the importance of syntactic processing to text comprehension. They assigned ninety-two 12 to 17-year-olds with LD to four different

groups. Participants read paragraphs either as they appeared in a textbook, with no changes, or with syntactic changes, semantic changes, or with both syntactic and semantic changes. Syntactic changes included things like clarifying the pronoun antecedent by eliminating relative clause sentences. For example, the sentence, “The strongest push came from Senator Douglas, who wanted the transcontinental railroad to begin in Chicago” was changed to “Senator Douglas wanted the transcontinental railroad to begin in Chicago.” The authors point out that this change also had the benefit of eliminating the idiom, “strongest push.” In some cases, a noun, even a redundant one, was substituted for an indefinite pronoun. Semantic modifications included things like reducing the number of multimeaning words: “minister” was changed to “ambassador.” Students read their assigned story and were asked ten questions that were devised based on informal reading inventories for assessing silent reading comprehension. The posttest contained one inference question, one vocabulary question, and eight factual questions. It was found that semantic changes alone did not significantly improve comprehension, but syntactic changes did, as did the semantic and syntactic changes together. The authors assert that, while texts cannot always be modified to suit the readers, training in the linguistic complexities identified in their study could enhance comprehension.

Taking a different approach, Tunmer, Nesdale, and Wright (1987) employed a reading level match design to demonstrate that syntactic awareness is related to beginning reading achievement. They matched good, younger (second-grade) readers with poor, older (fourth-grade) readers on four measures of reading ability—real word recognition, pseudo-word naming, reading fluency, and reading comprehension—and verbal intelligence. Syntactic awareness was assessed on two tasks: an oral correction task and

an oral cloze task. The good, second-grade readers performed significantly better on both the oral correction task and the oral cloze task than the poor fourth-grade readers.

Correlations between the two syntactic awareness tasks were highly significant for both groups, indicating that they measure a common underlying skill. That skill is the ability to reflect upon and manipulate the internal structure of sentences. These findings mean that the older, poor readers had a deficit in their syntactic awareness which may be a factor possibly limiting their reading comprehension.

Gaux and Gombert (1999) confirmed these results with a series of assessments administered to 83 French-speaking sixth-graders. Participants were given seven morphemic and syntactic assessments: repetition (repeating an orally presented sentence), judgment (judging the grammaticality of sentences), localization (identifying mistakes in a sentence), correction (repeating the sentence correctly), explanation (explaining a mistake), replication (reproducing two correct sentences with a corrected error) and identification of function (finding and circling a phrase or word with the same function as a target). Four reading tasks—including tests of decoding ability and comprehension ability—were administered. IQ, memory span, and oral comprehension were controlled factors. A regression analysis revealed that the more explicit the syntactic awareness tasks were, the more frequent and greater their contribution to reading performance. When the participants were divided into three groups—good, average, and poor readers—and their syntactic performance was compared, the poor comprehenders were found to suffer from a syntactic awareness deficit. The authors conclude that the link between syntactic tasks such as order inversion and reading comprehension could be due

to a failure to make inferences, since inferences are necessary to work out the possible meanings of scrambled sentences (Oakhill, 1994).

Morphemic and syntactic awareness might also facilitate word processing and the construction of grammatical structures (Gaux & Gombert, 1999, p.185). Fowler (1988) however, has argued that while judgment tasks directly test syntactic awareness, correction tasks require superior working memory. According to Fowler, good and poor comprehenders perform similarly on judgment tasks, but not on correction tasks, due to this memory load. His conclusion is that processing complexity, rather than structural complexity is a better diagnostic tool. As counterevidence, Gaux and Gombert (1999) point out there was no discrepancy between the poor readers' judgment and correction task scores in their study.

Finally, evidence for the relationship between syntactic awareness and comprehension has been suggested by results of training studies teaching the use of context. Developing new vocabulary words through context is hypothesized to rely, in part, on syntactic awareness in that decoding alone is not enough to determine which of several sounds a letter may represent (Nagy & Scott, 2000). Kuhn and Stahl (1998) reviewed these training studies and found that relatively brief interventions result in an overall increase in children's ability to learn new words. Moreover, although their studies were not strictly about syntactic awareness, protocols of students attempting to infer meanings of novel words from context in studies such as that of Werner and Kaplan (1952) and McKeown (1985) demonstrate the failure of low-ability readers to make syntactically appropriate inferences when faced with unfamiliar vocabulary. One way this is tested is by presenting students with a context sentence containing an artificial

word and asking the participants to fill in an acceptable real word. For example, in McKeown's (1985) study, a high-ability group child was asked to define the word *depcro* in the sentence, "The worried rider couldn't control the *depcro* horse." Her response was, *scared*. She reasoned that "somebody could have shot a bullet or something, made it scared, and he couldn't control it." In contrast, a low-ability group child chose the word *size* to substitute for *linbad* in the sentence, "After working on the problem, the group was ready to make a *linbad*." His reasoning showed that he chose *size* because the supporting context had been about making a dress and it was possible that the dressmakers did not know what size dress to make. By choosing a word of the wrong part of speech, the participant demonstrated that awareness of form class (syntactic awareness) is an important factor in vocabulary learning.

Finally, two training studies have shown that learning to reflect deliberately on language generally improves reading comprehension. Kennedy and Weener (1973) experimented with four groups of 20 third-graders each. One group learned to replace crucial words in an auditory cloze task. Another experimental group was trained in the cloze task with the sentences presented visually. One control group read orally from short stories and had decoding errors corrected; another control group stayed in the classroom. It was found that the experimental reading group did significantly better than the other groups on a reading comprehension subtest. They hypothesized that the cloze procedure training caused students to pay attention to units larger than the word.

Similarly, Weaver (1979) used a sentence anagram task and word-grouping strategy in her training study. Sixteen third-graders were taught to systematically arrange words into phrases and then to arrange the phrases into sentences. They formed word

groups by first identifying the verb and then determining how the rest of the words in the sentence supported the verb. When compared to 15 third-graders in a no-treatment control group equated for sex and reading level, it was found that sentence organization skills were trainable and that this skill transferred to reading comprehension as measured by a combination of scores on four separate tests.

Layton, Robinson, and Lawson (1998), however, have raised concerns about the procedures used in these two studies. It is not clear that the training tasks of Kennedy and Weener (1973) or Weaver (1979) increased syntactic awareness without also affecting the reader's semantic abilities. Layton, Robinson, and Lawson (1998) have proposed a four step theory of the development of metasyntactic awareness, based on Gombert's (1992) levels of general metalinguistic awareness. Level I is tacit knowledge of syntactic rules. While not in itself awareness, the authors consider it a prerequisite for the following levels. Level II is characterized by an ability to employ fix-up strategies, or to "manipulate the structural features of language" (p.8). This is the level commonly assessed (e.g. Bowey, 1986; Tunmer, Nesdale, & Wright, 1987). Level III requires the ability to articulate that rules of syntax exist and to identify some of them. Level IV is adult-like skill, requiring the ability to reflect on one's knowledge and/or performance on syntactic tasks. To obtain evidence, these authors then trained thirty fourth-graders in either semantic or syntactic awareness. The syntactic training consisted of small group instruction in the use of context, the "jobs words do" (p.13), and cloze tasks. The semantic group discussed word meanings and how to generate them. A control group was asked to help the research by answering questions such as "What is the easiest thing, and what do you like most about reading?" It was found that the syntactic training group was

successful at all four levels and their performance was stable over the course of two posttests, but the group did not show greater gains on a general reading ability test than the semantic training group. Layton, Robinson, and Lawson (1998) admit, however, that it is possible that the measures of reading performance used were not sufficiently sensitive to detect changes that may have occurred. In addition, they too may have suffered from the possible effects of low power, with only 30 total participants. However, whether Kennedy and Weener (1973) and Weaver (1979)'s interventions were purely syntactic in nature, or included semantic training as well, there is no question that their research demonstrates clearly the power of metalinguistic training.

Therefore, we can see from the research reviewed above that it is not entirely clear yet which aspects of metalinguistic awareness are most crucial to the reading process. Future research will test the importance of different aspects of linguistic awareness to different elements of reading. What *is* clear from this review is that there is no shortage of evidence to demonstrate the importance of being able to manipulate words and sentences independently of their contextual meaning.

Understanding ambiguity

Another body of research has demonstrated a relation between ambiguity detection and text comprehension. More than one type of ambiguity has been identified and studied. This section will review the types of ambiguity identified by linguists, as well as the role that the ability to disentangle ambiguity plays in reading development.

Lexical ambiguities are words with more than one meaning. For example, the sentence, "The man held the pipe" has a lexical ambiguity in the word *pipe*. It is unclear without further context whether the pipe in question is a smoker's pipe or a plumber's

pipe. An example of a sentence with a structural ambiguity is “The chicken is ready to eat.” Although none of the words contained in the sentence are ambiguous, there are two possible syntactic interpretations of the meaning: the chicken could be cooked and ready for someone else to eat, or it may be hungry and ready to eat its own dinner.

All ambiguity detection involves reprocessing. However, psycholinguists have shown that when lexical items are recognized, all meanings are activated in the lexicon, and the reader selects the most appropriate one (Swinney, 1979; Swinney, Prather, & Love, 2000). In contrast, only one structural analysis of a sentence is normally computed. To detect a structural ambiguity, the reader must reprocess while setting aside the preferred meaning (Cairns, 1999). This explains why detecting lexical ambiguities has been found to be substantially easier than detecting surface structure or deep structure ambiguities (Cairns, Waltzman, Schlisselberg, 2004; Shultz & Pilon, 1973). Rapid and efficient reprocessing of both lexical and structural items is necessary for competent reading comprehension.

This sort of linguistic reprocessing is unique in that it requires readers to construct two internally consistent representations for one sentence that conflict with one another (Yuill & Oakhill, 1991). The comprehension of non-literal language, such as joking, irony, sarcasm, and lying all rely on this ability. While there is a large body of research on all of these forms of language, the present research focuses primarily on jokes and riddles for their efficient and ecologically valid delivery of ambiguous language to beginning readers.

Cairns, Waltzman, and Schlisselberg’s (2004) investigation of first- and second-graders’ ability to detect the ambiguity in lexically and structurally ambiguous sentences

is one of the few studies to look at the relationship between reading and ambiguity detection. Participants in this study were presented with sentences such as *The children saw a bat lying by the fence* (lexical) or *The chicken was ready to eat* (structural) and were asked what they thought the sentence meant. Once the participant gave a preferred meaning, the investigators asked if the sentence could mean anything else. If the child was unable to produce another meaning, the investigator provided a series of prompts. It was found that first-graders did well with the lexical sentences but only the second-graders detected the structural ambiguities. Furthermore, ability to detect lexical ambiguities correlated highly with the results of the Slingerland (1977) test of reading readiness administered at the beginning of first grade, and with scores on the Gates-MacGinitie Reading Test (MacGinitie, MacGinitie, Maria, & Dreyer, 2000) subtests of comprehension, vocabulary, and decoding administered in second grade. In the third grade, both lexical and structural ambiguity detection correlated strongly with reading. Multiple regression analysis showed that lexical ambiguity detection when the children were pre-readers accounted for half the variance in the total second-grade reading scores and that second grade ambiguity detection was a strong predictor of third grade reading scores.

Cairns, Waltzman, and Schlisselberg (2004) hypothesized that ambiguity detection skill is a predictor of early reading ability. They determined that the ability to understand ambiguity is developmental, with eight-year-olds on the cusp of full metalinguistic awareness. The ability to detect lexical ambiguities develops in the first grade and is a strong predictor of second grade reading ability. Likewise, the ability to detect structural ambiguities (i.e., phrases in which the syntax lends itself to multiple

meanings) emerges in second grade and predicts third-grade reading ability. This finding implies that specific instruction in ambiguity, as well as how to use context clues to know when reprocessing is necessary, could lead to more efficient comprehension. In order to understand ambiguous words or sentences in text, the reader must understand that ambiguity exists and what ambiguity is, as well as how to regulate one's cognition about it.

Other studies that have looked at the relationship between ambiguity understanding and reading have used riddles as the vehicle for assessment. As mentioned earlier, riddles are fun and familiar short texts, which make them particularly appropriate for teaching and assessing young children. Many different types of riddle studies exist. For example, Fowles and Glanz (1977) and Yalisove (1978) investigated the development of riddle comprehension. Hirsh-Pasek, Gleitman, and Gleitman, (1978) and Shultz (1974) then looked at the relationship of riddle comprehension to age changes. Mahony and Mann (1992) related riddle solving ability to reading. Most recently, Yuill (1996, 1998) and Yuill and Oakhill (1991) treated riddle solving ability as an index of metalinguistic awareness.

Most riddles involve multiple meanings, but they vary in the level at which the ambiguity occurs. Several different classifications have been used (see Yalisove, 1978). Table 1 shows the categories and examples designed by Yuill (1996) in order of increasing size of the linguistic unit involved in the ambiguity.

Other researchers have used riddles that do not involve linguistic ambiguity to provide contrast. Yalisove (1978) calls these "absurd" riddles (e.g., *How do you get six elephants in a mini car? Three in the front and three in the back*). These riddles share the

Table 1

Yuill's (1996) Riddle Classification

<i>Riddle Type</i>	<i>Example</i>
Morphophonological (phonological manipulation by segmenting a word or by replacement with a similar sounding phoneme or morpheme)	Where do you take a hurt wasp? To the waspital.
Lexical (relies on the double meanings of words)	Why was the crab arrested? Because he kept pinching things.
Word compound (contains two words in a compound that need to be unpacked and rearranged)	Would you like a pocket calculator, sir? No thanks—I know how many pockets I've got.
Syntactic (ambiguity is in the syntactic structure)	How do you stop a fish from smelling? Cut off its nose.
Pragmatic (violates expectations about the speaker's intent)	Why do firemen wear red suspenders? To keep their pants up!
Metalinguistic (contrasts the use and the mention of terms for linguistic entities)	What word is loud, even when you say it softly? Loud.

same format as linguistic ambiguity riddles, even if they do not require the same reasoning processes. They are therefore useful as controls in many studies.

In one of the first studies to use riddles to look at the relationship between reading and the ability to comprehend riddles, Hirsh-Pasek, Gleitman, and Gleitman (1978) asked forty-eight children from grades 1-6 to rate the humor and explain the joke contained in a variety of linguistic riddles presented orally. Children's responses were scored on a five-point scale representing whether or not they got the joke immediately and explained both sides of the ambiguity. The authors predicted that riddles that rely on semantic properties (lexical and underlying structure riddles) would be easier for the participants to explain than riddles that turned on superficial properties (phonological and surface-structure ambiguities). Their theory was that it is more natural to focus on meaning than on syntactic form. Indeed, they found that the easiest riddles to explain were the ones containing ambiguities that turned on lexical interpretations or transparent semantic properties (e.g., *How can hunters in the woods best find their lost dogs? By putting their ears to a tree and listening to the bark*). Their subjects had comparatively more difficulty with phonological ambiguities (e.g., *If you put three ducks in a box, what do you have? A box of quackers*) and surface structure ambiguities (e.g., *Where would you go to see a man-eating fish? A seafood restaurant*), as well as what they termed deep structure ambiguities (e.g., *Will you join me in a bowl of soup? Do you think there's room for both of us?*). Hardest of all were morpheme boundary distortion riddles, either with (e.g., *Do you think that if I wash, my face will be clean? Let's soap for the best*) or without phonological distortions (e.g., *How do trains hear? Through their engine ears*). While

there was a main effect of grade and reading ability, no interactions were observed between these subject variables and any of the ambiguity types.

Hirsh-Pasek, Gleitman, and Gleitman (1978) interpret this as support for their theory that as general verbal talent increases, so does one's ability to access superficial aspects of language. This stands in juxtaposition not only to Tunmer and Bowey's (1984) theory that awareness of higher levels of language is related to later-developing, more sophisticated, reading skills, but also to the many studies that have since demonstrated the early importance of phonological awareness to decoding ability (e.g. Bryant and Bradley, 1985; Lundberg, Frost and Petersen, 1988). Hirsh-Pasek, Gleitman, and Gleitman (1978, p.103) state, "...humans find it easiest to access high-level (or fully processed) linguistic representations, and much more difficult to process lower-level (syntactic and phonological) representations." They also, however, point out that when phonological riddles are delivered orally the pronunciation is always biased, making the riddle either easier or harder for the listener to understand. In addition, the lack of interactions between reading ability and riddle type means that the poorer readers did not do significantly worse on the riddles with surface feature ambiguities. This is not in keeping with their prediction that skilled readers focus more easily on syntactic properties.

Another difficulty with Hirsh-Pasek, Gleitman, and Gleitman's (1978) study is that they interpreted "general verbal talent" to be grossly measured by reading ability. However, they did not measure the reading abilities of their participants, but rather just split them into good and poor readers based on the opinion of their classroom teachers. Yuill and Oakhill (1991) suggest that their "poor" readers were probably actually above

average in reading skill. Nevertheless, using their rough classification, Hirsh-Pasek, Gleitman, and Gleitman determined that, at every grade, the better readers were better able to solve the riddles. The poor readers performed at a level about a year behind the good readers in explaining the ambiguities.

As indicated, this study has severe limitations. The relative comprehension and decoding skills of the participants are unknown, so the results of this study cannot be generalized to a broader population. Sample size is an issue as well: with only eight participants at each grade level, it is ambitious to make generalizations about the development of reading skills in relation to the development of linguistic skills. Ryan and Ledger (1984) point out that in Hirsh-Pasek, Gleitman, and Gleitman's (1978) study as well as other cross-sectional studies (for example, Shultz & Pilon, 1973), individual differences within age groups were greater than differences across ages. This indicates a need for longitudinal research to follow the development of individuals.

Mahony and Mann (1992, 1998), on the other hand, predicted that second graders' reading skill would correlate with the ability to choose the correct punch line for phoneme/morpheme riddles (i.e., those that turn on sublexical units), but not for control riddles with no morphophonological ambiguity. They tested 48 second-grade children on their word reading, decoding, and receptive vocabulary, and their ability to resolve riddles. Ten riddles specifically dependent on phoneme/morpheme boundary confusion and ten control riddles (composed of ambiguities at the lexical, logical, and syntactic levels) were selected and alternative punch lines were created. The child's task was to determine the "answer that made the riddle the funniest" (p.174). They found that reading ability—as measured with the Word Identification and Word Attack subtests of

the Woodcock Reading Mastery Tests, Form A (Woodcock, 1973)—was significantly related to the ability to resolve riddles which manipulated phonemes and bound morphemes, but not control riddles. They also found that, for their participants, IQ (as measured by the Peabody Picture Vocabulary Test [Dunn & Dunn, 1997]) was significantly related to performance on both types of riddles, but IQ was not significantly related to reading ability. Thus, their findings support the idea that facility with riddles is a metalinguistic skill separate from, but related to, reading ability, and that this relationship is not a result of IQ.

Yuill (1996) agreed that phonological/morphological riddles should be related to decoding accuracy rather than comprehension, whereas riddles operating at a lexical level and above should be related more closely to comprehension. No previous studies had related riddle appreciation to reading comprehension (Yuill, 1996, p.199). Yuill therefore administered a task of riddle disambiguation to a group of 29 8-11 year-olds along with the Neale Analysis of Reading Ability (Neale, 1966). The Neale Analysis provided the researcher with age-related scores separately for word reading accuracy and comprehension. Accuracy scores were based on the number of word reading errors. Comprehension scores were based on comprehension questions—both literal and inferential—answered from memory after reading a short story. The riddles were categorized according to the size of the linguistic units involved, as defined by Tunmer and Bowey's (1984) model. Children's recall was measured for each type of riddle by asking the children to rate the funniness of a riddle, to re-tell the riddle, and then to explain the riddle back to the experimenter immediately after having heard the riddle presented orally. Although she does not give means, Yuill reports that the explication

task was too difficult. She therefore focused on the recall scores. It was found that riddle recall was significantly related to reading comprehension in two of five riddle categories: (1) word-compound (e.g., *Would you like to buy a pocket calculator, sir? No thanks, I know how many pockets I've got*), (2) metalinguistic (e.g., *What word is loud even when you say it soft? Loud*).

One problem with this study involved the use of riddle recall as the outcome measure assessing riddle understanding. Yuill (1998) defends recall as a measure of comprehension because the wording of a particular word or phrase is often crucial to maintaining the integrity of the interpretation. While an explication task would have been preferable, riddle explanation is a difficult task even for adults, and is inappropriate for young children without supporting context.

In a follow-up study, Yuill (1998) administered a riddle task in which a similar group of participants were given questions with two different punch lines, the intended one and one that did not resolve the incongruity, and they were instructed to choose the best answer for each riddle. In this follow-up study, she also introduced two additional categories of riddles: morphophonological riddles (e.g., *What kind of keys are funny? Monkeys*) and syntactic riddles (e.g., *What animal can jump higher than a house? All animals, houses can't jump*). She expected that morphophonological riddles would tap awareness of low-level aspects of language and thus correlate with reading accuracy but not with comprehension, in contrast to word-compound and metalinguistic riddles that were found to correlate with reading comprehension in the previous study. No prediction was made for syntactic riddles. Indeed, Yuill found significant positive correlations between the riddle solving task and reading comprehension for all riddle types except the

morphophonological riddles. Furthermore, the morphophonological riddles were the only riddle type to correlate positively with reading accuracy.

Finally, in a third study, Yuill (1998) demonstrated that metalinguistic training improves the reading comprehension of seven- and eight-year-olds. Eighteen participants were given seven treatment sessions of about 30 minutes each over a period of seven consecutive weeks. They made up jokes, played games with ambiguous communications and clue-constructions, and read abstract and/or ambiguous texts (although the exact procedures are not made explicit in the published version of the study). These students were matched on reading comprehension ability to eighteen students forming a control group who read silly stories and played phonemic awareness games. It was found that both groups increased their comprehension scores, but the children given ambiguity training showed a significantly greater mean increase in comprehension age by about six months. These results seem to indicate that it is possible to teach comprehension skills to younger struggling readers in a short and engaging teaching program based on flexibility with words and attention to syntax.

An alternative explanation for these results, however, is that comprehension monitoring acted as an intervening variable. If the ambiguity training taught participants to evaluate their understanding of the text, that may have been enough to improve their regulation as well. Yuill (1998) did not administer a comprehension monitoring measure, so there is no way to pinpoint whether it was facility with language or better comprehension monitoring (or a combination of the two) that accounted for the increase in reading comprehension scores. In addition, the ambiguity training in Yuill's (1998) intervention is not well described, so it is impossible to know whether the tutoring she

experimented with would translate easily to the classroom. We do know that her participants were chosen to be children whose reading comprehension lagged significantly behind their decoding skill, and as such, do not represent an average classroom. The present research project clarifies and extends Yuill's work by teaching ambiguity and language flexibility to third graders who range in reading ability. It is expected that this training will increase the participants' awareness and knowledge of ambiguity, comprehension monitoring, and in turn, overall reading comprehension.

The research studies presented above demonstrate the value of being able to detect and comprehend ambiguity in text. As we have seen, many of these studies utilized children's riddles in their assessment and/or training materials, due to the children's comfort with the form, the convenience of being able to find so many ambiguous phrases in one place, and the inherent motivation behind such a fun tool.

Comprehension monitoring

The final intersecting piece of the research puzzle is comprehension monitoring. Reviewed here are studies on when readers fail to notice that they have not understood what they are reading.

Beginning in the 1970s, researchers began to discover that reading comprehension is an active process. Researchers such as Markman (1981), Gagne (1977) and Weinstein (1978) studied the awareness that readers have of their own comprehension processes during reading. Isakson and Miller (1976) found that good comprehenders are much more likely to notice that a mis-decoded word does not make sense in the given context than poor comprehenders are. Markman (1981) noted that children often fail to ask for more information when presented with ambiguous messages (Cosgrove & Patterson,

1978; Ironsmith & Whitehurst, 1978). She hypothesized that this could be due to the lack of sufficient processing space for realizing that more information is required, for noting alternative interpretations, and for holding tentative decisions in memory until a determination can be made on whether or not reprocessing is necessary.

The rise of interest in metacognitive processes gave birth to a new area of research within reading: the role of metacognition in text comprehension, that is, comprehension monitoring. Interest in comprehension monitoring was increased by the seemingly contradictory findings that very young children are capable of monitoring their listening comprehension (Baker, 1984; Patterson et al., 1981; Revelle, Wellman, & Karabenick, 1985), whereas adults often do *not* monitor their reading comprehension (Baker, 1989; Baker & Anderson, 1982; Rinehart & Platt, 1984).

Although comprehension monitoring has long been assumed to be a crucial element of effective reading (Dewey, 1910; Thorndike, 1917), Baker (1979) and Markman (1979) were among the first researchers to operationalize and investigate this process. Markman (1981) wrote about the phenomenon of letting one's mind wander while reading. She reports that sometimes, while entirely engrossed in a daydream, she will read every word, for pages on end, without comprehension *or knowledge of her own incomprehension*. This is an extreme example of failure to monitor comprehension. Markman went on to uncover evidence for children's failure to monitor their comprehension. She found that young children sometimes claim to understand instructions even though they contain glaring omissions (1977). In addition, older elementary students' judge essays to be perfectly comprehensible despite blatant inconsistencies (1979). In follow-up studies, other researchers have also found that

children deem instructions which contain contradictions to be perfectly comprehensible (Wagoner, 1983).

While many researchers have asserted that poor monitoring is an element of poor reading comprehension (Cornoldi & Oakhill, 1996), the directionality of the relationship is in dispute (Cornoldi, 1991). Poor monitoring may be a result of poor comprehension rather than the cause. Ruffman (1996) looked into the possible reasons for monitoring failure.

Ruffman theorized that there are six possible sources of difficulty behind comprehension monitoring. He labeled these difficulties conceptual, metacognitive, low-confidence knowledge or nascent knowledge, information-processing limitations, a predisposition to derive a conclusion, and an absence of constructive processing (p.33). The first in this list, the idea that the child may have a different concept of what constitutes problematic text is quickly ruled out, as there is plenty of evidence that even very young children can identify empirically inconsistent statements (Pea, 1980), and—by the age of six—have an understanding of logical consistency (Ruffman, 1993). The other five are not so easily dispatched.

Ruffman defines metacognitive difficulty as not being able to distinguish between guessing at the meaning of a text (because it is unclear) and really knowing the meaning (because it is clear). Perner (1991) provided evidence that children have some understanding of the guess/know distinction by 6-years-old, however this contrasts with the evidence showing that even adults have difficulty with comprehension monitoring tasks. Ruffman suggests that this could be evidence either that our understanding of when

we do not truly understand something continues to develop very slowly, or it could mean that some factor other than the metacognitive contributes to comprehension monitoring.

Low confidence or nascent knowledge, or reticence to commit to an answer, is third in Ruffman's list of possible sources of difficulty in comprehension monitoring. Ruffman suggests that low confidence knowledge is most often conflated with other sources of difficulty in the tasks researchers use to measure comprehension monitoring. Often comprehension monitoring tasks are designed with a short text containing an anomaly. Children then read or listen to the text and are asked if it makes sense. But Ruffman argued that a better test would be a two-story task in which one story contained a logical inconsistency. The participants would then read and compare both stories, then judge, "Which story made more sense?" In this way the participant is forced to acknowledge there is a problem and commit to an answer, making it more likely that the child with nascent or low-confidence knowledge will succeed. However, when this theory was tested (Ruffman, 1995) with forty 8-year-olds randomly assigned to either a standard one-story condition or a two-story condition, no evidence was found for low-confidence knowledge.

Ruffman (1995) next tested the theory that working memory impeded the success of the students in the two-story condition from his experiment above. As the fourth possible difficulty impeding comprehension monitoring, many researchers have theorized that information-processing limitations are to blame for comprehension monitoring failure (e.g., Vosniadiadou et al., 1988). Evidence for this includes Yuill and Oakhill's finding that poor comprehenders are significantly worse than good comprehenders at finding inconsistencies in text, when the inconsistencies were spread out, rather than

adjacent (Yuill & Oakhill, 1991). Other researchers have also found that without training, children are better at detecting empirical (external) inconsistencies than logical (internal) inconsistencies (Baker, 1984; Vosniadiadou et al., 1988). Ruffman argues that logical inconsistencies are identified by comparing propositions within the text, whereas empirical inconsistencies are evaluated against known fact. In any event, when Ruffman replicated the experiment described above, this time with sixty-nine 7- and 8-year-olds randomly assigned to one-story or two-story conditions, and controlled the number of exposures students received to each line in each story, no effect was found for either low-confidence knowledge or working memory.

Ruffman, Billins, and White (1995) investigated children's predisposition to derive a conclusion. They tested the hypothesis that, when given a number of possible conclusions, children may answer questions about their own knowledge by checking to see if they have an answer. The conclusion first reached or stressed in the text becomes the preferred conclusion, even though there may be another possible meaning. Therefore, in this experiment, children read two stories (consistent and inconsistent) and were asked not only whether or not the story made sense, but whether or not the story would make sense to a doll named Sarah. The children performed significantly better when asked about the doll than about themselves. The results also revealed that children over-attributed knowledge to themselves, but not to the doll. The children said yes, the story made sense to both consistent and inconsistent stories because they checked their knowledge and were able to infer a conclusion.

Next, three versions of a story were prepared: consistent, inconsistent, and no information (about the disputed fact in the inconsistent story). Participants received

either the consistent and inconsistent stories or the inconsistent and no-information stories. It was hypothesized that if children determine inconsistent texts to be clear because they provide information (too much information) to reach a conclusion, they should find it easier to report the uninformative text as unclear than to identify the inconsistent text as unclear. Indeed, the results and the children's attributions confirmed this hypothesis. In some cases, the children explicitly said that they chose the inconsistent story as the clearer one simply because it was more informative ("Cuz this one says more," p.57).

Finally, Ruffman (1996) considered whether or not children engage in constructive processing in the absence of comprehension monitoring. To do this, he first examined the reading times for each sentence of each story across the experiments. His results indicated that children spent more time reading lines that contradicted earlier lines than they did reading consistent lines. Even a cautious interpretation of this data led Ruffman to the conclusion that failure to spot inconsistency is not necessarily due to a failure to engage in constructive processing. This is at odds with Markman's (1977, 1979) idea that children's failure to spot inconsistencies or monitor their comprehension is due to a failure to engage in constructive processing. This finding is also at odds with the findings of Oakhill (1982) and others, who have suggested that good comprehenders are more likely to engage in constructive processing than poor comprehenders. Ruffman suggests that perhaps both good and poor comprehenders are equally involved in building mental models of the text, but that good comprehenders are just more likely to succeed.

Other researchers have also found evidence for a relation between the meaning readers construct from text and their comprehension monitoring (e.g. Baker & Anderson,

1982; Baker & Zimlin, 1989; Kinnunen & Vaurus, 1995; Zabucky & Ratner, 1986, 1992). Kinnunen and Vaurus (1995) reviewed these studies. Baker and Anderson (1982) investigated the comprehension monitoring ability of college students by presenting expository texts sentence-by-sentence on a computer screen for comprehension and later recall by the participants. Comprehension monitoring was measured in reading time and look-backs. Passages included internal inconsistencies related to either a main point or detail in the passage. After reading, subjects were asked to answer multiple choice questions and to identify the inconsistencies. Analyses of reading times and look-backs showed evidence of monitoring for both main ideas and details. Baker and Anderson (1982) interpret this to mean that mature readers monitor their comprehension as they read. Kinnunen and Vaurus (1995), however, criticized the study for using only college students as participants. They point out that college students are necessarily competent in constructing integrated meanings for large text units. It would have been more illuminating to compare the performance of subgroups representing different levels of reading skill. In addition, Baker and Anderson (1982) do not report the relation between the recall measures and monitoring.

Zabucky and Ratner (1986, 1992), on the other hand, employed third and sixth graders in one study (1986) and good and poor sixth grade readers in another study (1992) in their investigation of children's comprehension monitoring. The children were videotaped reading stories with internal inconsistencies, one sentence at a time. Location of the inconsistencies was varied to be either close to or far from the context-giving sentences. Participants' comprehension evaluation was measured on-line with reading times and verbal reports. Comprehension regulation was examined by evaluating look-

backs during reading. It was found that all of the children read inconsistencies more slowly, and that good readers (1992) and sixth graders (1986) were more likely to look back at inconsistencies during reading, to report these inconsistencies, and to recall the text after reading. Zabrocky and Ratner (1986, 1992) did not examine the effect of far or close context conditions, except to point out that look-backs can constitute either a re-examining of the context sentence (from earlier in the text), or a re-reading of the incongruent sentence.

Kinnunen and Vaurus (1995) wondered if longer reading times in the far text condition could be explained by text features, such as topic shifts, rather than as evidence of comprehension regulation. After examining of these studies, Kinnunen and Vaurus (1995) performed their own experiment. They compared the level of comprehension monitoring and level of text-integration in low- and high-achieving fourth graders, and examined the effect of training on the low-achieving participants. Eye-movements were recorded to evaluate comprehension monitoring so that readers could concentrate on reading for comprehension while the researchers recorded both reading time and look-backs. Inconsistencies in the text included word level, sentence level, and internal discrepancies. Comprehension was measured with main idea questions and a summarization task. Clear differences were found between the reading of low- and high-achieving participants. The high-achievers read faster, performed better on the comprehension questions, summarized more clearly, and monitored at all levels of text processing. By contrast, the low-achievers monitored their comprehension only at the lower (lexical) levels.

The low-achieving children were then split into three groups. Two groups were administered a 16-week program devoted to cognitive strategies and metacognitive skills. One of these training groups also received socio-emotional coping therapy to address their feelings of inadequacy, which the authors note has been shown to coincide with non-task-oriented behavior. These children were shown video samples of negative emotional responses, avoidance tendencies, and self-deprecatory causal attributions, and they discussed these samples extensively. The third group served as controls.

In the strategy training, the researchers used direct explanation, scaffolding, self-guided and cooperative practicing, and articulation and reflection to teach comprehension skills. Text-processing training first occurred at two levels. At the propositional text level students were taught to activate relevant prior knowledge, to detect difficult words, and to reason about their meanings from context. At the global text level, students were taught deletion and selection rules for determining main ideas, as well as condensing and connecting ideas within the text. After that, students were taught integration strategies, such as summarizing a complex text. Kinnunen and Vaurus point out that this means that few comprehension monitoring skills were directly taught. Instead, comprehension monitoring skills were embedded in the training of text comprehension skills. Therefore, they label the posttest tasks of comprehension monitoring, which involve identifying external and internal inconsistencies, *transfer* measures.

In analyses of training effects it was found that the high-achievers still performed significantly better than the trained low-achievers. The researchers did, however, find a positive effect for training. Low-achievers from the trained groups spent more time

reading internally inconsistent lines, indicating that the training raised their level of comprehension monitoring to the local level. By contrast, the untrained low-achievers slowed their reading rate only for lexical inconsistencies in the text. Participants from the trained group also wrote more concise summaries, indicating that they were building better connections between sentences. The only measure on which they did not improve is the number of look-backs, suggesting they still did not have the automatic strategies the high-achievers did for overcoming internal inconsistencies.

Baker and Zimlin (1989) also conducted an intervention study, based on the comprehension monitoring standards developed by Baker (1985). Baker (1985) posited that comprehension monitoring is a multidimensional process. She distinguished seven criteria which control a reader's evaluation and regulation of understanding:

- 1) *lexical* standards for monitoring the comprehension of words;
- 2) *syntactic* standards for monitoring the grammaticality of a phrase or sentence;
- 3) *internal consistency* standards for checking that ideas in the text are not contradictory;
- 4) *external consistency* standards for making sure that world knowledge is not violated
- 5) *propositional cohesiveness* standards for checking that there is a cohesive relationship among propositions;
- 6) *structural cohesiveness* for evaluating the thematic compatibility of ideas; and
- 7) *informational completeness* standards for checking that the text provides all the necessary information for comprehension.

Baker and Zimlin (1989) conceptualized lexical, external consistency, and propositional cohesiveness as “microstructure standards” (defined as requiring local attention to words and sentences) and structural cohesiveness, internal consistency, and informational completeness as “macrostructure standards” (which require global text integration). Eighty fourth-graders identified as average and above-average readers on the basis of their California Achievement Test scores were divided into three groups. One group was given instruction and practice in the microstructure standards. One group learned about and practiced macrostructure standards. The third received no instruction. After a 30-minute training session, the participants read a 250-word passage and were asked to identify parts that were hard to understand. The test passage contained problems corresponding to all six of the standards. Experimental condition participants were told explicitly that the passages contained types of problems beyond those they had just learned about. Control group participants were instructed to find the problems, but not told what types of problems may be there. All participants were retested 2-3 weeks later. It was found that students in the training conditions found more problems than students in the control condition; that both training groups found more non-instructed problems than the control group; and that, although better readers outperformed average readers, the effect of instruction was equivalent. These findings indicate that most children can benefit from training in metacomprehension, and that training in some skills generalizes into more effective monitoring in other areas as well.

Baker and Zimlin (1989) also included a supplementary hypothesis that training in the use of macrostructure standards would result in greater generalization than training in the use of microstructure standards. Their reasoning was that text macrostructure is

derived from the microstructure (see Kintsch & van Dijk, 1978), and so successful use of macrostructure standards—i.e. successfully considering the relation between ideas in a text—would require as a prerequisite paying attention to individual propositions and sentences, thus making students in the macrostructure instruction group more sensitive to problems at the microstructure level as well. They found, however, no evidence to support this hypothesis. In fact, there was no generalization of macrostructure training to lexical items. Students from the control group found about the same number of lexical problems in the text as the macrostructure training group, whereas students from the microstructure training group found significantly more.

This finding is congruent with the direction of the causal relation hypothesized in the present study: attention to lexical ambiguity is proposed to be part of the metalinguistic awareness that is applied in comprehension monitoring, which in turn strengthens reading comprehension. Rather than being a prerequisite for the use of macrostructure standards, microstructure standards are complementary and directly influence comprehension. What a reader needs to know to monitor comprehension is different at the micro- and macro-levels. Teaching cognitive strategies and providing modeling and practice time for planning and evaluation may be the steps that are necessary for helping students to regulate their global awareness. At the micro-level, however, it may be that explicit instruction in how language works is just as helpful as global strategies. In terms of memory performance, it is logical that practice and instruction in microstructure analyses are as important as macrostructure analysis. Greater lexical access and easier sentence parsing should lead to more rapid and automatic word identification. This automatic processing requires less memory, since the

reader does not have to consciously decode or puzzle out the meanings of the sentence, thus enabling attention to the inferential processing necessary for comprehension monitoring (Perfetti & McCutchen, 1986).

On the other hand, Hacker (1997) found a significant interaction between instructions to use different standards of text evaluation and the number of corrections made to text. He used an error detection paradigm to investigate the comprehension monitoring of adolescents as they searched three times through a text. With a cross-sectional experimental design, Hacker asked 315 seventh-, ninth-, and eleventh-graders of low, average, and high reading ability (based on standardized test scores) to find the errors planted in a text. Each time the participants searched the text they were given more explicit instructions. The text consisted of seven expository paragraphs estimated to be at the seventh-to-eighth grade reading level and containing 24 errors, 8 spelling, 8 grammar, and 8 internal consistency errors (as described by Baker, 1985). In Search One, no specific instructions were given—students were merely told to read the essay and then to find and correct errors in it. Hacker found that all of the readers failed to spontaneously monitor many of the problems, but that monitoring at lexical, syntactic, and semantic levels all increased with age and reading ability. In Search Two, Hacker divided the students up into four different instructional groups. One group was instructed to again “find the errors;” another group was told that some errors affect the meaning of the essay (what Hacker called semantic errors), and was given examples of those type of errors; a third group was given examples of spelling and grammatical errors (what Hacker called lexical and syntactic errors); and the instructions for a fourth group made reference to “lexical, syntactic, and semantic errors” (p. 7). The results of Search Two

revealed that there was an asymmetry in the effect of instruction such that instructing students to look for meaning errors increased their detection of meaning errors with no reduction in their detection of spelling or grammatical errors, but that instructing students to look for spelling or grammatical errors increased their detection for those kind of errors while decreasing their ability to find meaning errors. In Search Three, the experimenter provided monitoring assistance by pointing to the lines that contained errors. The students were able to correct these errors once pointed out, showing that some students had the knowledge necessary to monitor text, and yet still failed to apply it on their own.

Hacker (1997) concluded that students identify more errors overall when taught to look for semantic errors than when taught to look for lexical and syntactic errors. This would seem to support Baker and Zimlin's (1989) hypothesis that teaching macrostructure monitoring is more important than teaching microstructure monitoring. However, the problem here is in classifying what constitutes lexical and syntactic awareness. Misspellings are certainly errors on a lexical level, and grammatical mistakes can be syntactic. Hacker does not consider, however, lexical and syntactic misinterpretations that can change the reader's comprehension of text. For example, all of the ambiguous sentences from Cairns, Waltzman, and Schlisselberg (2004) contained lexical or syntactic elements that change the meaning of the sentence. In the sentence *The woman looked over the fence*, the meaning changes based on the grammatical roles assigned to the words. Thus the reader needs to be syntactically aware to comprehend the sentence. This kind of syntactic awareness can change the reader's comprehension of the text, and thus is considered more appropriate for Hacker's macrostructure category than

the microstructure. Hacker, however, attempted to distinguish two categories of instruction given to experimental groups: those who were told to look for macrostructure errors that changed the meaning of the text, and those who were asked to find microstructure errors that did not change the meaning of the macrostructure. He did not allow for the possibility that lexical and syntactic errors can be classified as macrostructure.

However, the ability to flexibly assign grammatical roles to words in a text is a form of syntactic awareness. As we have seen from the studies reviewed in the previous chapters, syntactic awareness and the ability to monitor one's understanding are the necessary ingredients in detecting and repairing the meaning interpretations of ambiguous text. Therefore, it stands to reason that understanding ambiguity would have a significant impact on a student's reading comprehension.

The ability to recognize mistakes in a text is an important skill in reading because it indicates that the reader recognizes when he/she has not comprehended. Many researchers have shown that both adults and children fail to monitor their comprehension, although they have the ability to do so. Ruffman (1996) identified some possible sources of this difficulty. Other researchers (i.e., Baker & Zimlin, 1989; Kinnunen & Vaurus, 1995) have had success in improving comprehension monitoring through strategy training programs.

Chapter 3: Preliminary Studies

Two previous studies conducted by the author lend support to the hypothesis that training in ambiguity detection can lead to increased reading comprehension.

Correlational Study

A study conducted by the experimenter (Zipke, 2005) provided correlational support for the possibility that metalinguistic training may aid reading comprehension. Based on the work of Yuill (1998), in which 8-11-year-olds' riddle solving ability was correlated with their reading comprehension ability, it was hypothesized that sixth- and seventh-graders who scored higher on standardized tests of reading comprehension and vocabulary would also be better able to solve linguistic riddles with punch lines that were dependent on homonyms or the structural arrangement of words within the sentence and to recognize sentences with lexical and structural ambiguities.

To test this hypothesis, one-hundred sixth- and seventh-graders were administered the Gates–MacGinitie Reading Comprehension and Vocabulary subtests, Level 7/9 (MacGinitie et al., 2000), a riddle punch line task, and a test of ambiguous sentence recognition. For the riddle punch line task, 25 riddles were typed on a sheet of paper in random order, with two punch lines each, one which solved the riddle, and one which did not. Participants were asked to read the items silently to themselves, and then to choose the answer that made the question a riddle. Each child received a score indicating the number of riddles identified correctly out of 25 maximum. For the ambiguous sentence recognition task, 40 sentences (i.e., 10 with deep structure ambiguities, 10 with surface structure ambiguities, 10 with lexical ambiguities, and 10 foil sentences) were printed in

random order with a 1 and a 2 printed after each sentence. Students were asked to read the sentences silently to themselves, decide if the sentence had one meaning or two, and to circle the corresponding number. Intercorrelations between riddle solving, ambiguous sentence recognition, and performance on the vocabulary and reading comprehension subtests were examined: significant correlations were found among all four measures.

A multiple regression revealed that vocabulary accounted for 77% of the variance in the reading comprehension scores, a finding consistent with that of Stanovich (1992), and others. Riddle solving also accounted for significant unique variance in the reading comprehension scores, adding about 3% to the total variance after vocabulary had been entered. This was interpreted as evidence that metalinguistic awareness is an ability that is separate from general linguistic intelligence and that contributes to reading comprehension.

Ambiguous sentence recognition did not contribute to the variance in the reading comprehension scores in this study. Moreover, the ambiguous sentence performance and riddle solving tasks were not found to measure the same thing, as the correlation between the two was only .24. Very likely the ambiguous sentence recognition task did not contribute to the variance in reading comprehension scores as a result of problems with the task itself. The ambiguous sentence task was not as interesting for the participants—in contrast to the riddles task, which had an inherent humorous pay-off that resulted when the puzzle was solved. So it may have been that, when faced with forty boring sentences, good as well as poor readers simply did not put as much care into reading and thinking about each sentence as they did into each riddle. Because the ambiguous sentence task was administered first, fatigue was not a factor in this consideration.

Riddles have a form that is familiar and fun for children. It is unlikely that the participants were familiar with the particular riddles chosen for the task: scores revealed no floor or ceiling effects. And all of the riddles chosen for the riddle task were carefully screened for their reliance on homonyms, structural ambiguities, and deep structure confusions. Therefore, the riddle task is believed to have been an accurate measure of the participants' metalinguistic abilities.

In agreement with the findings of Yuill (1998), this study provided strong evidence that there is a relationship between reading comprehension and a flexible understanding of language. The riddle task assessed the participants' ability to flexibly test out multiple meanings of homonyms and garden path ambiguities, rearrange the syntax in some sentences, and reflect on the deep structure meaning of some of the others in order to successfully choose the punch line which made the question a riddle. In this way, the riddle task measured the larger construct of metalinguistic ability. As expected, the results of this task correlated highly with the reading comprehension scores.

Furthermore, as the findings of the multiple regression show, metalinguistic awareness accounts for significant unique variance in the reading comprehension scores, after vocabulary is statistically controlled. It is not surprising that vocabulary emerged here as the single best predictor of reading comprehension; this finding concurs with that of Stanovich (1992), and others. In addition, it is well known that vocabulary correlates highly with IQ scores. Therefore, we can consider the vocabulary scores as a rough proxy for a measure of general verbal intelligence and conclude that metalinguistic awareness is not simply a function of intelligence at this age. This is convergent with the experimenter's perception that it is not necessarily the "smartest" children in class who

most enjoy riddle books; understanding and enjoying word play stems from a kind of intelligence not measured directly by traditional school grades.

Training Study

However compelling, the results of this study are correlational. In order to draw definitive causal conclusions about the effect of metalinguistic awareness training on reading comprehension, an experimental study employing a control group was planned for the dissertation research. The purpose of the study was to use materials that lend themselves easily to classroom adaptation to train low-SES New York City public school students in metalinguistic awareness.

In a pilot to the dissertation study, ten third-graders were taught that words, phrases, and sentences can have more than one meaning, as well as how to determine which meaning the author intended by focusing on the context. It was expected that the participants who received the metalinguistic training would improve in their metalinguistic awareness and in their reading comprehension more than the control group, who received traditional reading comprehension instruction. All of the participants were chosen from across a broad range of comprehension abilities because it was believed that this sort of instruction would be beneficial to students at all ability levels.

Potential participants were screened for minimum decoding ability of no more than one grade level behind expected. Those children who scored satisfactorily on the decoding tests were then administered a standardized reading comprehension test. All of the participants were matched into pairs according to these scores and randomly assigned to experimental or control groups.

The screening for potential participants consisted of two tests: The word

identification and non-word identification reading subtests of the Woodcock Reading Mastery test (Woodcock, 1987). In addition, the Gates-MacGinitie Test of Reading Comprehension, Level III (MacGinitie et al., 2000) was administered in different forms at pre- and post-test along with an ambiguous sentence task devised by Cairns, Waltzman, and Schlisselberg (2004) and modified by the experimenter. For the ambiguous sentence task, the experimenter read an ambiguous sentence aloud to the participant and asked the participant to explain as many meanings as possible. Participants received credit for understanding the ambiguity only if they were able to successfully articulate both meanings of the sentence. Lastly, the Peabody Picture Vocabulary test (Dunn & Dunn, 1997) was administered as a post-test only.

Instruction was carried out individually, in 6 sessions of 30-45 minutes each over a period of about 4-6 weeks. The metalinguistic training consisted of an active learning program utilizing riddles and authentic texts. A control group, made up of 10 participants matched for decoding and comprehension ability, received lessons without a metalinguistic component.

Metalinguistic instruction built from the ground up, to work with words first, then sentences, then text. In the first training session, homonyms were defined and brainstormed, and then the student practiced finding the homonyms among the words in a word tub.

In the second session, the student was shown ambiguous sentences with two different illustrations to demonstrate the different meanings. The child then listened to a structurally ambiguous sentence, such as *The ball was found by the kitten*, and practiced

building both meaning representations with Colorforms (a type of reusable, manipulable sticker).

In the third session, participants first practiced identifying *who* did *what* to *whom* in the sentences from the previous week. Then a detective game was introduced, in which the child was presented with a story three or four sentences long in which one sentence was syntactically ambiguous. The child was challenged to find the sentence with more than one meaning and explain it to the investigator.

The fourth session began with the detective game. Next, idioms were introduced and defined. The instructor supplied idioms, and the participant represented both the literal (unintended) and the idiomatic meaning in drawings. At the end of the session participants read text requiring inferences and answered comprehension questions based on those inferences. For example, a paragraph about a boy who peddles home never specifically says that he traveled by bicycle, but the comprehension question asks which mode of transportation he employed.

In session five riddles were introduced. The experimenter and participant shared favorite riddles, then wrote riddles on a topic of particular interest to the student.

Finally, in the last session, the participant read aloud from *Amelia Bedelia and the Surprise Shower* by Peggy Parish (Parish, 1979). The child was instructed to stop reading every time there was a sentence with more than one meaning, in order to explain how Amelia will interpret the sentence, as well as how the student knew the intended meaning of the sentence.

Results of the pilot study were encouraging and suggestive. The training proved successful in improving comprehension of ambiguous sentences. A one-way analysis of

variance was significant, $F(1,18) = 10.15, p < .01$, with the children who received metalinguistic training gaining a mean of 1.9 sentences from pretest to posttest and the control group losing ground with a 1.1 mean decrease. This indicates that the treatment was effective in teaching ambiguity detection and explication. Moreover, experimental participants improved more than control participants in their standardized reading comprehension scores, although not significantly so. A one-way analysis of variance was conducted to determine if the experimental tutoring had a significant effect on the participants' reading comprehension gain. Results of that analysis indicated that the null hypothesis cannot be rejected, $F(1, 18) = .182, p > .05$. The mean increase for the experimental group was 2.2 items as opposed to 1.5 items for the control group.

Finally, the correlation between performance on the Peabody Picture Vocabulary test and the students' gain in reading comprehension was found to be insignificant. This was interpreted to mean that the instruction was age-appropriate and that general verbal ability was not a factor.

These results were interpreted to mean that ambiguity training, of the sort described by Yuill (1998), is effective with third graders in a short school-based program. Furthermore, the positive trend in the reading comprehension score gains led the researcher to the cautious conclusion that this type of metalinguistic awareness may contribute to increased reading comprehension. An expanded and strengthened experiment based on the one described above was therefore planned for the dissertation study.

Chapter 4: Dissertation Study Rationale and Design

The purpose of the dissertation study was to utilize the same design as the pilot study to examine the effect of metalinguistic training on third graders' reading comprehension. The dissertation study involved a modification and extension of the pilot study. Several possible shortcomings of the pilot study were identified and rectified. For example, one reason the pilot results fell short of significance may be the test itself. The Gates-MacGinitie subtest of reading comprehension, Level 3 (MacGinitie et al., 2000) proved too easy for the test group. An examination of scores on the pre- and posttests revealed at least one perfect score. The test consisted of 39 items, but the mean score on the pretest was 29.6 with a standard deviation of 4.1; the posttest mean was 31.5, with a standard deviation of 3.8. This lack of variability may have hindered the detection of a difference. The instruction itself, on the other hand, was deemed to be age-appropriate and general verbal ability not a factor in the analysis, as the correlation between scores on the Peabody Picture Vocabulary test and the students' gain in reading comprehension was not significant.

The study described herein remedied the ceiling problem by replacing the Gates-MacGinitie reading comprehension subtest, Level 3 (MacGinitie et al., 2000) with Level 4, and by adding the Woodcock Reading Mastery Tests-Revised, subtest of passage comprehension (Woodcock, 1987) as an additional standardized reading comprehension outcome measure. The Woodcock passage comprehension subtest (Woodcock, 1987) is a better measure for this study because it is a cloze task. Instead of reading a passage and answering questions about it, the reader is required to read a sentence silently and then

supply the missing word. While the reader must comprehend the text in order to supply an appropriate word, he/she also needs to be syntactically aware in order to successfully use context clues to determine the part of speech needed. Cloze tasks like this one have been criticized, however, for not asking the reader to integrate information between sentences (Harris & Sipay, 1990). The two measures in combination were intended to provide a better assessment of the participant's reading comprehension than either test alone (Shankweiler et al., 1999).

In addition, some of the training methods were modified to strengthen the intervention. Some of the training tasks were changed or omitted. It was determined that the idioms and the hidden inferences tasks were not directly germane to the desired learning. The 3W chart (i.e., Who did What to Whom) was introduced as soon as ambiguous sentences were introduced. The final task in the training, the book reading, was strengthened to include active participation of the students. Instead of merely reading the *Amelia Bedelia* books, some of the pages were removed so that the reader had to predict Amelia's alternative interpretation of the words on the page before the experimenter revealed the double meaning. After reading *Amelia Bedelia's Family Album* (Parish, 1988), participants wrote their own *Amelia Bedelia*-type story based on the double meanings of the names of various professions. These changes reduced the duration of training from six sessions to four.

It was expected that with a larger sample size and stronger methods and materials than those used in the pilot study, the results would demonstrate the efficacy of teaching metalinguistic awareness to third-graders. Specifically, it was expected that the participants who received the ambiguity training would show significantly greater

improvement from pretest to posttest on the measures of metalinguistic awareness, comprehension monitoring, and reading comprehension than would the participants who were matched for comprehension ability and assigned to the traditional reading comprehension group.

The significance of this study cannot be overstated. Teachers and parents all over the world labor every day to find new and innovative ways to help children with their reading difficulties. Some of these difficulties are easily classified and/or remedied, but others are not. And the longer these difficulties go unaddressed, the more detrimental they become to effective school learning.

Reading comprehension is an enormously complicated construct. Thus far, research has focused on selected aspects that have been shown to have a large effect on text comprehension (e.g., vocabulary, background knowledge, decoding ability, etc.), but it is still possible that there are other instructional components that contribute to the understanding that a reader constructs from text.

Based on the research studies reviewed here, there is good reason to believe that awareness of syntactic and semantic cues can be facilitated in such a way as to aid students in their comprehension monitoring. Furthermore, there is every reason to believe that this will improve their reading comprehension. In sum, this is a new direction for comprehension instruction—one that may assist in unraveling the mysteries of reading and reading disability.

Chapter 5: Method

Participants

Participants were selected from two public schools in Brooklyn, New York. The population of the schools was multicultural, with 63% of the students qualifying for a free lunch, indicating that the schools were located in a lower-class neighborhood. Approximately 5% of students in the schools are English Language Learners. Participation in this study was limited to students who were fluent English speakers and who did not have individualized education programs (IEPs). All of the children who returned permission slips (see Appendix A for the consent form and Appendix B for the child's oral consent script) were screened for word reading ability, decoding, and receptive vocabulary knowledge. From students who scored at a second grade-equivalent level or above on a standardized word reading test, participants were chosen randomly. After pretesting, participants were matched on their reading comprehension scores, and mates were randomly assigned to either the experimental training group or the control group.

The sample consisted of 46 third graders, 23 assigned to the training group and 23 assigned to the control group. The mean age, gender and ethnic makeup of each group are given in Table 2.

Materials and Procedures

Tasks were administered in the order listed below. Screening was conducted in one session of about 30 minutes. Participants were then randomly selected from those whose word reading ability was no more than one year behind their expected grade level.

Table 2

*Characteristics and Mean Performance of the Metalinguistic Treatment and Control**Groups on the Screening Tests and Pretests*

	Treatment Group	Control Group	<i>t</i> stat (<i>df</i> = 22)
Age (in months)	103.22 (3.42)	104.65 (4.01)	
Gender	11 male, 12 female	14 male, 9 female	
Ethnicity			
White	14	11	
African-American	6	7	
Latin-American	2	3	
Asian-American	1	2	
WRMT-R Word ID ^a	59.52 (7.61)	62.30 (13.36)	-1.12 <i>ns</i>
Grade Equivalent	3.7	3.9	
WRMT-R Word Attack ^a	23.35 (6.42)	24.30 (9.83)	-.50 <i>ns</i>
Grade Equivalent	2.8	2.9	
PPVT ^b	101.83 (12.53)	102.13 (16.81)	-.09 <i>ns</i>
WRMT-R Comp ^b	100.09 (7.12)	101.09 (11.44)	-.55 <i>ns</i>
Grade Equivalent	3.0	3.0	
GMRT4 ^b	47.13 (17.31)	48.48 (18.83)	-.33 <i>ns</i>
Homonyms (10 max)	2.78 (2.28)	2.83 (2.53)	-.08 <i>ns</i>
Ambiguous Sents (8 max)	1.00 (1.35)	1.87 (1.96)	-1.97 <i>ns</i>
Riddles (5 max)	2.61 (1.23)	2.87 (1.49)	-.81 <i>ns</i>

Table 2, con't

Characteristics and Mean Performance of the Metalinguistic Treatment and Control Groups on the Screening Tests and Pretests

	Treatment Group	Control Group	<i>t</i> stat (<i>df</i> = 22)
Heteronyms (10 max)	3.87 (2.12)	4.30 (2.82)	-.63 <i>ns</i>
Self-corrections (percentage corrected)	.144 (.214)	.239 (.233)	-1.23 <i>ns</i>
Substitutions in reading graded text	18.09 (10.86)	13.52 (5.51)	1.86 <i>ns</i>
Anomaly detection (10 max)	6.13 (2.88)	6.04 (3.28)	.11 <i>ns</i>

* $p < .05$; *ns* = not significant

Note. There were 23 students in each group. PPVT is the Peabody Picture Vocabulary Test. WRMT-R Comp is the Woodcock Reading Mastery Test—Revised, reading comprehension subtest. GMRT4 is the Gates-MacGinitie Test of Reading Comprehension, Level 4.

^aRaw scores.

^bStandard scores. Mean for norming sample on WRMT-R Comp is 100. Mean NCE score for norming sample is 50.

Eight tasks were administered as pretests and then repeated as posttests following training. In all cases, the posttest was identical to the pretest except for the items. All tasks and training sessions were administered individually, except for the Gates-MacGinitie Test of Reading Comprehension (MacGinitie, et al., 2000), which was administered to small groups of 3-4 children at a time. The pretesting and posttesting sessions were conducted individually and consisted of the first 7 tasks. Each session lasted about 45 minutes. Training was likewise conducted individually, in sessions of about 45 minutes duration, occurring once per week over the course of four weeks. The instructor was familiar to staff and students at the schools and sessions were conducted in the school resource room or library.

Screening tests:

- 1) *Word Identification subtest, from the Woodcock Reading Mastery Tests-Revised (WRMT-R)* (Woodcock, 1987). Participants read aloud from a list of words of increasing difficulty. The WRMT-R word identification subtest has a publisher-reported split half reliability of .97.
- 2) *Word Attack subtest, from the Woodcock Reading Mastery Tests-Revised* (Woodcock, 1987). Participants read aloud from a list of non-words of increasing length and complexity. The WRMT-R word attack subtest has a publisher-reported split half reliability of .91.
- 3) *Peabody Picture Vocabulary Test (PPVT), third edition* (Dunn & Dunn, 1997). A measure of receptive vocabulary. Participants listen to a vocabulary word presented orally by the experimenter and choose one of four black-and-white line drawings that best depict the meaning of the word. This vocabulary score was

considered a correlate of verbal intelligence. The PPVT-III has a publisher-reported internal consistency reliability alpha of .95.

Pretests and posttests:

- 1) *Homonym definition test*: Metalinguistic awareness. In order to assess participants' metalinguistic awareness at the word level, the experimenter read each participant a list of 10 homonyms and asked the participant to describe as many meanings for each word as possible. After each definition, the participant was asked, "Does ___ mean anything else?" Students often responded to this prompt with additional functions of the term rather than additional definitions. For example, the first definition given for the homonym "sink" was often "a place where you wash your hands." When asked if "sink" could mean anything else, many participants responded with, "a place where you wash dishes." This response was noted but not corrected, and the next item was presented. The number of words for which participants suggested more than one distinct meaning was scored (for example, for "sink" another distinct meaning would be the verb form: "to fall down in water"). The alpha reliability for this test was .82. (See Appendix C for a complete list of all of the homonyms given as prompts in the pretest and posttest and the instructions given participants.)
- 2) *Ambiguous sentence detection task*: Metalinguistic awareness. In order to assess participants' metalinguistic awareness at the sentence level, participants listened to 8 ambiguous sentences one at a time, and were asked to explain as many alternative meanings as possible to the experimenter (as in Cairns, Waltzman, &

Schlisselberg , 2004). Prior to testing participants heard the sentence, *The chicken is ready to eat* and were shown the following pictures:



Figure 1



Figure 2

Participants were told that one meaning of the sentence is that the chicken is hungry and ready to eat his dinner (represented in Figure 1). Another meaning is that the chicken is cooked and ready for other people to eat (represented in Figure 2). A second practice example, *They talked about the problem with the teacher*, was presented and discussed. Then the eight test sentences were presented and the participant was instructed to give the first meaning thought of for the sentence and then to try to think of another meaning. Participant responses varied, but most students either responded immediately with the second meaning or with, “I don’t know.” Alpha reliability for this test was .83. (See Appendix D for a complete list of the sentences used as prompts for this task and the instructions given participants.)

- 3) *Riddle resolution test*: Metalinguistic awareness. This task was adapted from the metalinguistic awareness assessment used by Zipke (2005), in which participants silently read questions followed by two punch lines and selected the punch line that turned the question into a riddle. This time, the experimenter read each of the five questions followed by two punch lines aloud to each participant. Participants were told that a riddle is “a puzzling question that ends with an answer that surprises you and usually makes you laugh. The question and answer make a riddle when the same words have two different meanings.” They were further

instructed that only one of the punch lines contained an ambiguity that gave the riddle multiple meanings and that their job was to choose the punch line that makes the question into a riddle. For example, for the question, *Why did the skeleton go to the movies by himself?* the participant was expected to choose the punch line, *He had no body to go with him* rather than, *He was lonely*. The number correct was recorded. Aside from spontaneously giggling, students did not comment on this task other than to choose a punch line, as instructed. Alpha reliability for this test was .65. (See Appendix E for a complete list of riddles and the instructions given participants.)

- 4) *Heteronym pronunciation*: Comprehension monitoring. In order to assess participants' comprehension monitoring at the word level, participants read aloud ten sentences containing two instances of heteronyms, that is, words that are spelled the same but pronounced differently (for example, *I lowered my bow and arrow and took a bow*). Participants were given no instructions other than to please read each sentence aloud. The number of sentences in which both heteronyms were pronounced correctly was scored. Alpha reliability for this test was .81. (See Appendix F for a complete list of the sentences used as prompts and the instructions given participants.)
- 5) *Miscue self-correction*: Comprehension monitoring. In order to assess participants' comprehension monitoring at the word level, but within an authentic reading context, participants read aloud from a series of graded paragraphs about animals, history, science, and fairy tales that increased in difficulty until their errors exceeded 10% of the text. Participants were given no instructions or

rationale other than to please read aloud. The ratio of self-corrections to miscues within passages that were read at or above 90% accuracy was calculated. (See Appendix G for the graded paragraphs and instructions given participants.)

- 6) *Anomaly detection*: Comprehension monitoring. In order to assess participants' comprehension monitoring on the story level, participants read aloud from familiar stories (*The Three Little Pigs* and *The Three Bears*) containing micro- and macro-structure anomalies. Baker and Zimlin (1989) conceptualized "microstructure standards" as those requiring local attention to words, sentences, structural cohesiveness, and internal consistency, and "macrostructure standards" as those which require global text integration. The participant was instructed to stop at pre-determined points in the text to explain his/her thinking. The instructions were: "Tell me what you are thinking. For example, if you see something you think is wrong, or doesn't make sense, or if what you are reading reminds you of something else, or if you think you know what will happen next, tell me about it." Participants who did not stop at the pre-determined points or who merely repeated text were prompted with, "Tell me what you are thinking." The number of anomalies for which the participant correctly identified the problem or attempted to resolve the anomaly was recorded. For most of the prompts, the expected response was cut-and-dried: for example, most of the participants protested immediately that the name Cinderella does not belong in the story of Goldilocks and the Three Bears. In other instances, however, the child attempted to make sense of the anomalous information and fit it into the story instead of merely protesting that it did not belong. For example, after one boy

read the section of *The Three Bears* that says, “Then she decided to sit down. First she sat in Papa Bear’s chair, but it was too hard. Then she sat in Mama Bear’s chair, but it was green,” he did not protest that this was wrong or use his knowledge of the story to claim that it should say that Mama Bear’s chair is supposed to be too soft, as most of the children did. Instead, he talked at length about how it shouldn’t matter to Goldilocks if the chair is green, but how it is possible that she really hates green... for example, maybe she had a bad experience with something green and now she is scared to sit in green chairs, etc. This was scored as a correct response because the participant noticed the anomaly and attempted to integrate the information into what he was learning from the rest of the text. Student readings and responses were taped, transcribed, and assessed by two independent raters. It was explained to the raters that for all ten items in each story, if the child mentioned the anomaly in his/her thinking aloud it should be scored positively, whether or not the child ultimately decided that the item made sense in the text or not. So, for example, when a child responded to the anomaly about the “big bad bear” (instead of “wolf”) with the think aloud response, “Oh, I get it...he’s saying he’s a bear because he’s *as hungry as* a bear,” his response was scored positively. Raters disagreed over participant comments that mentioned the anomaly, but made irrelevant observations. For example, after the prompt, “When upset, most bears watch TV,” one student said, “They said most bears watch TV, but there’s not a bunch of bears. It’s not really *most* cuz’ [sic] there’s only three bears.” Interrater reliability was 90%. Alpha reliability for this test was .92. The pretest story was *The Three Little Pigs* and the posttest story

was *The Three Bears*. (See Appendix H for the texts and instructions given participants.)

- 7) *Passage Comprehension subtest, from the Woodcock Reading Mastery Tests-Revised* (Woodcock, 1987, 1998). The format of this reading comprehension test involved the cloze procedure. Participants read text silently and provided missing words that were cued by a blank space with short phrases and increased in length up to five or six sentence paragraphs. There were 68 possible items in all, but most students read 32 to 48 items before completing the publisher-specified ceiling criteria of six consecutive failed responses that end with the last item on an easel page. The experimenter recorded the word supplied by the participant, but did not provide feedback. If no word was supplied within 30 seconds, the experimenter urged the participant to try the next item. According to the publisher's manual the split-half reliability of the Woodcock Reading Mastery Test subtest of passage comprehension is .92
- 8) *Gates–MacGinitie Test of Reading Comprehension, Level 4 (GMRT4)* (MacGinitie et al., 2000): Paragraph reading comprehension. The GMRT4 is a timed test of reading comprehension in which participants read eleven passages each of approximately three paragraphs in length and answer multiple choice comprehension questions. The test was group-administered to two or three students at a time. According to the publisher's manual, the KR-20 reliability of the GMRT4 is 0.91.

Procedure

Training: Metalinguistic Awareness of Multiple Meanings

Participants received four individually administered training sessions on words and sentences with multiple meanings (See Appendix I for script). Instruction began at the word level and grew gradually to the reading of authentic texts. With the use of active learning methods and commonly found classroom materials, new concepts were presented and demonstrated, then handed over to the participant for guided practice. Specifically, the four sessions consisted of the following procedures:

Session 1: Multiple Word Meanings. The first session was devoted to introductions, explaining the purpose of the tutoring, and determining that words can have more than one meaning. The instructor defined “homonym” for the children and then together they brainstormed words that the children knew to have more than one meaning (cold, poor, etc.) The instructor tested the participants by giving multiple definitions for a few words and asking the children if the definitions contrast. For example, some were intentionally bad examples (i.e. “a ball can refer to either a baseball or a basketball”), which led to a discussion about similar traits. When the children each produced at least three words on their own, the instructor introduced the word tile tub.

The word tile tub contained about 40 words in all parts of speech. Twenty-two of the words had multiple meanings. Three of the homonyms from the tub had already been discussed in the teaching portion above. The children were challenged to find ambiguous words hiding in the tub. Two primary strategies were taught for identifying homonyms. The first strategy was to consider whether the word could fit another form class (e.g. if it’s a noun, can it also be a verb?). The second strategy was to generate synonyms to uncover potentially different definitions and then to compare the synonyms (e.g. the baseball and basketball are both still toys). The session ended when participants had

categorized all of the words in the tub. The number of homonyms that the participant correctly identified and defined on his/her own was recorded. As they left, the children were asked to think of more words that have more than one meaning for the next session.

Session 2: Multiple Sentence Meanings. First, the participants were read ambiguous sentences and shown two different illustrations. The instructor discussed the fact that both illustrations represent the same sentence, and explained how the two meanings of the sentence differ. For example, the children were first given the sentence, “The dog chased the man on a bike,” and shown one illustration featuring a dog running after a bicycle with a man atop it, and an alternate illustration featuring a man running away from a bicycle-riding dog. In this teaching portion of the session the experimenter demonstrated alternative meanings for the children and did not ask the participants to attempt any definitions. This demonstration continued with seven different sentences.

Next, the instructor read the children sentences with structural ambiguities. The children were challenged to represent one meaning of the sentence with Colorforms (a type of reusable, manipulable sticker). The instructor corrected as necessary. After the children represented one meaning, they were asked to represent another meaning. For example, for the sentence, “The ball was found by the kitten” the participants were given a scene and expected to hide the ball near the kitten and then to introduce another character that finds the ball. To display the alternate meaning, the participants were expected to place the ball anywhere on the scene and then have the kitten find the ball.

Beginning with the second sentence, the instructor introduced a poster board chart labeled with three large Ws. It was explained to the children that the three Ws stand for *who does what to whom*. As they progressed through the sentences, the experimenter

charted the three Ws step-by-step for one meaning of the sentence and then for the other meaning.

Example: “The ball was found by the kitten.”

WHO: the kitten
(did) WHAT: found
(to) WHOM: the ball

WHO: Bob the Builder
(did) WHAT: found
(to) WHOM: the ball (next to the kitten)

The number of sentences for which the participant was able to independently illustrate both of the meanings was recorded. When they left, the participants were asked to think of more sentences that have more than one meaning for the next session.

Session 3: Riddles. This session was devoted to riddles. The instructor introduced the concept of lexical riddles first: “Why do spiders like baseball? They’re good at catching flies.” Then she introduced structural riddles: “Where can you see a man eating fish? A seafood restaurant.” She explained the answers to the participant with words, the 3W chart, and manipulables (if possible). The participants were asked to volunteer any riddles they knew.

Loosely following the procedures laid out in Bernstein (1979) for writing riddles with children, the instructor and participant discussed a topic of particular interest to the participant for which s/he knew many specific terms. A list of such words and phrases was brainstormed and then examined to see if any words on the list had multiple meanings that could fit into the context of a riddle. The child was then scaffolded in their attempt to produce riddles. In this way, the children created a riddle book devoted to the chosen subject. The number of riddles independently created by the participant was

recorded. When they left, the participant was asked to think of more riddles for the next session.

Session 4: Text Reading. In this last session, the participants graduated to book reading. First the children read *Amelia Bedelia and the Surprise Shower* (Parish, 1979). In the Amelia Bedelia series, Amelia frequently misunderstands directions: she does things like “trim” the steak with ribbons and lace or “make a jelly roll” by prodding jelly along the floor. The children were instructed to stop every time they came to a sentence with more than one meaning. They then explained to the instructor how Amelia understood the sentence, as well as how the sentence is meant to be understood. Furthermore, the instructor asked the child to explain how he/she knew which one was the intended meaning.

Next, the children began reading *Amelia Bedelia’s Family Album* (Parish, 1988). In this book, Amelia introduces her family members one by one, with their name and profession, and the interpretation Mr. or Mrs. Rogers puts on Amelia’s explanation. The next page, however, contains Amelia’s alternate interpretation (for example, a “boxer” to Amelia is someone who puts things in boxes). After the first few examples, the instructor covered Amelia’s interpretations in the participants’ books so that they were required to speculate on what Amelia would say. Each child was given the chance to decipher Amelia’s point of view. Finally, the participants wrote their own entries into Amelia’s family album with professions brainstormed by the participant and experimenter. The number of professions for which the participant correctly deduced Amelia’s interpretation as well as the number of entries to the book written independently by the participant were recorded.

At the beginning of each new session, participants were asked if they remembered what they had been taught in the previous week, and if they had come up with any new homonyms/sentences/riddles on their own. These answers were elicited primarily as a warm-up to the current session. Responses were neither recorded nor analyzed.

Control Group Training: Book Reading and Discussion

Control group participants also met with the experimenter once per week, but for shorter periods of time (often 10-15 minutes) to read aloud from *Mouse Soup* (Lobel, 1977), a storybook at a reading level equivalent to that of the *Amelia Bedelia* books read by the participants in the metalinguistic training condition. (See Appendix J for script.) Correct decoding of all words in the text was stressed.

Prior to reading, the investigator initiated a discussion with the participant about the components of the book: author, illustrator, title page, table of contents, etc. Following the reading, the discussion focused on stylistic elements of the written text such as plot, point of view, setting, and imagery. The purpose of this control group was to address the possibility of Hawthorne effects. The control group did not receive any training in metalinguistic awareness, but became familiar with the experimenter and believed they were part of a larger program of reading instruction.

Design and Statistical Analyses

A pretest/posttest experimental design with matched pairs and random assignment was used. The experimental group received metalinguistic training and the control group received standard reading comprehension instruction. Paired sample t-tests were used to verify that the treatment and control groups did not differ significantly on any of the screening and pretest measures. Two-way ANOVAs were used to assess the effectiveness

of metalinguistic training on outcome measures. The independent variables were type of training (metalinguistic awareness vs. standard reading comprehension) and time of test (pretest vs. posttest). The latter variable was a repeated measure. To assess whether treatment effects held across test items as well as students, the proportions of students in the two groups succeeding on each test item were compared.

Chapter 6: Results

Characteristics of Participants

The participants ranged greatly in ability, thus reflecting the distribution in an average classroom. Characteristics of the participants are shown in Table 2. There were 25 boys and 21 girls in the sample. The mean age was 8 years and 8 months. At the outset of the project, participant word reading ability ranged from 36 (a second grade equivalency) to 90 (a grade equivalency of 10.2). Despite a couple of unusually high-scoring students, the mean was 61, with most students scoring in the third grade range. Similarly, at pretest, reading comprehension ability varied from 17 (a grade equivalency of 1.8) to 45 (a grade equivalency of 6.9) on the Woodcock Reading Mastery Tests-Revised, subtest of passage comprehension (Woodcock, 1987, 1998). Reading comprehension scores varied from 12 (a grade equivalency of 2.0) to 46 (a grade equivalency of 10.4) on the Gates–MacGinitie Test of Reading Comprehension, Level 4 (MacGinitie et al., 2000). Standardized vocabulary scores ranged from 61 to 136, with a mean of 102. Thus, third graders in the sample were performing on average at expected levels when compared to national norming samples on these tests.

Participants were matched on the basis of their reading comprehension scores and mates were randomly assigned to groups. Paired-sample *t*-tests confirmed that performance of the two groups did not differ on any of the pretest measures (see Table 2). Because scores on the self-correction text reading measure represented a ratio of corrections to substitutions, another paired sample *t*-test was performed to ensure that the control and experimental groups had substituted an equal number of words on the pretest. Results are presented in Table 2.

Effectiveness of Metalinguistic Training: Comparison of Experimental and Control Groups

Two objectives of the study were to determine whether metalinguistic awareness is trainable in a simple and fun school-based program, and whether training in metalinguistic awareness enhances reading comprehension. To answer these questions, two-way ANOVAs were conducted to assess improvement of the two groups on the outcome measures from pretest to posttest. The independent variables were treatment (experimental vs. control), and time of test (pretest vs. posttest). The latter was a repeated measure. This analysis was applied to each of the eight dependent variables, which included three metalinguistic awareness measures, three measures of comprehension monitoring, and two standardized reading comprehension assessments. Table 3 reports the full results of these analyses.

Significant main effects of test time as well as an interaction between time and treatment were found for two of the three metalinguistic tasks: homonym and ambiguous sentence definition. From Table 3, it is apparent that students who received metalinguistic training improved much more from pretest to posttest than control students did. The trained students produced more multiple meanings of homonyms as well as more multiple meanings of ambiguous sentences than control students did. Both types of responses had been taught and practiced during the training. It is important to note, however, that four types of ambiguous sentences were practiced during training but none of these specific types were tested on the posttest. This indicates broader transfer from training to test as compared to the homonym training/assessment, which was much more

Table 3

Mean Performance of Metalinguistic Treatment Group and Control Group on Posttests and Test Statistics

	Treatment	Control		<i>F</i> stat (1, 44)	<i>d</i>
Homonyms (10 max)			[G]	2.01 <i>ns</i>	
Pretest	2.78 (2.28)	2.83 (2.53)	[T]	15.39**	
Posttest	5.22 (2.65)	3.35 (2.66)	[G*T]	6.44*	.70
Gain	2.44	.52			
Ambiguous Sentences (8 max)			[G]	.20 <i>ns</i>	
Pretest	1.00 (1.35)	1.87 (1.96)	[T]	22.92**	
Posttest	2.91 (1.70)	1.61 (1.90)	[G*T]	39.68**	.72
Gain	1.91	-0.26			
Riddles (5 max)			[G]	.06 <i>ns</i>	
Pretest	2.61 (1.23)	2.87 (1.49)	[T]	.03 <i>ns</i>	
Posttest	2.74 (1.51)	2.65 (1.70)	[G*T]	.47 <i>ns</i>	.06
Gain	.13	-.22			
Heteronyms (10 max)			[G]	.13 <i>ns</i>	
Pretest	3.87 (2.12)	4.30 (2.82)	[T]	.07 <i>ns</i>	
Posttest	4.13 (1.91)	4.17 (2.76)	[G*T]	.58 <i>ns</i>	-.01
Gain	.26	-.13			

Table 3, con't

Mean Performance of Metalinguistic Treatment Group and Control Group on Posttests and Test Statistics

	Treatment	Control		<i>F</i> stat (1, 44)	<i>d</i>
Self-corrections			[G]	.88 <i>ns</i>	
Pretest	.144 (.214)	.239 (.232)	[T]	3.92 <i>ns</i>	
Posttest	.257 (.193)	.260 (.207)	[G*T]	1.85 <i>ns</i>	
Gain	.11	.02			-.01
Anomaly detection (10 max)			[G]	1.02 <i>ns</i>	
Pretest	6.13 (2.88)	6.04 (3.28)	[T]	2.52 <i>ns</i>	
Posttest	7.43 (2.95)	5.70 (3.75)	[G*T]	7.50**	.51
Gain	1.30	-0.34			
GMRT4			[G]	.00 <i>ns</i>	
Pretest	47.13 (17.31)	48.48 (18.83)	[T]	1.29 <i>ns</i>	
Posttest	46.17 (16.59)	45.30 (23.69)	[G*T]	.37 <i>ns</i>	
Gain	-.96	-3.18			.04
Woodcock Passage			[G]	1.3 <i>ns</i>	
Pretest	100.09 (7.12)	101.09 (11.44)	[T]	1.42 <i>ns</i>	
Posttest	105.22 (8.32)	97.96 (11.33)	[G*T]	24.23**	.73
Gain	5.13	-3.13			

p* < .05; *p* < .01; *ns* = not statistically significant.

Note. Cohen's *d* = $M_1 - M_2 / \sigma_{\text{pooled}}$ where $\sigma_{\text{pooled}} = \sqrt{[(\sigma_1^2 + \sigma_2^2) / 2]}$. Effect sizes were calculated on posttest means.

closely matched. Taken together, these results show that the training was effective in teaching students to think flexibly about the meanings of words and sentences.

No significant effects were found for riddle solving. In this task, students had to choose which of two punch lines solved each of five riddles. Chance performance was 2.5 items correct. From Table 3 it is apparent that scores hovered around chance level. To test this, 2.5 was subtracted from each student's riddle score and then these difference scores were compared to zero. For the experimental group on pretest, $t(22) = .42, p > .05$; for the control group, $t(22) = 1.19, p > .05$. For the experimental group on posttest, $t(22) = .76, p > .05$; for the control group, $t(22) = .43, p > .05$. These results show that performance was no higher than chance. In other words, students may have been unable to solve the riddles and so guessed at the answers.

Among the comprehension monitoring tasks, a significant interaction between test and condition was found for the anomaly detection think aloud measure. From Table 3, it is apparent that trained students detected more of the anomalies in the story than control students on the posttest but not on the pretest. Whereas scores of trained students increased, scores of control students decreased slightly. Macro and micro items were also analyzed separately for the posttest items. There were four micro items and six macro items for this task, so each child's micro score was divided by four, and each child's macro score was divided by six to find percentages of the total. A matched pair t-test was calculated in order to compare the trained students' performance on each type of item with that of the control group on the posttest. No effect was found for either type of item. For the micro items the mean was .07 with a standard deviation of .43, $t(22) = .73, p >$

.05; for the macro items the mean was -.04 with a standard deviation of .51, $t(22) = -.41$, $p > .05$. These results could be due to the restricted number of items in each analysis.

On two other comprehension monitoring tasks, no significant main effects or interactions were detected. Trained students were not more likely than control students to self-correct their word substitution errors as they read text passages orally. Also, in the heteronym task, trained students did not differ in the extent that they produced two different pronunciations of the same spelling as they read the words in sentences. One possible reason for the lack of differences is that these types of comprehension monitoring skills were not taught or practiced during training.

To summarize, the metalinguistic training group showed much greater improvement from pretest to posttest than the control group on two measures of metalinguistic awareness and one measure of comprehension monitoring. In fact, the control group showed a loss rather than a gain on the ambiguous sentence measure and the anomaly detection measure. These findings indicate that metalinguistic training was successful in improving students' ability to identify multiple meanings of homonyms and ambiguous sentences, and to detect anomalies embedded in stories as they read them.

The impact of training on reading comprehension was also examined. A significant main effect of test time as well as a significant interaction between test and treatment was detected on the WRMT-R passage comprehension subtest (Woodcock, 1987, 1998), but not on the GMRT4 reading comprehension test (MacGinitie et al., 2000). Students who received metalinguistic training outperformed control students in completing cloze test items to indicate comprehension of text, but the two groups did not differ on a silent reading task that required reading passages and answering multiple-

choice questions. These findings support the hypothesis that metalinguistic training improves students' reading comprehension ability, at least as it is assessed by some measures.

The lack of a treatment effect on the GMRT4 was not surprising. Observation of participants as they completed this reading comprehension test raised doubt about its validity. Students were observed marking answers without reading the passages. Also, some were observed to look at the answer sheets of other students, despite remonstrations from the investigator. This behavior was not possible during administration of the WRMT-R, as that was an individually administered cloze task.

For each of the test items in the outcome tasks that showed significant effects of training, the proportion of participants in each group who got each item correct was examined to determine if the positive effects detected for participants also held across items. Results are shown in Tables 4, 5, and 6. For the tasks on which main effects had previously been found—homonym definition, ambiguous sentence recognition, and anomaly detection—the experimental group outperformed the control group on every posttest item, with the exception of one ambiguous sentence and one anomaly which were explained correctly by an equal number of experimental and control group participants. These findings show that almost all of the test items on these tasks were sensitive to the effects of metalinguistic training.

Response to Metalinguistic Awareness Training

Students' success during training was analyzed both qualitatively and quantitatively. First, each session was analyzed and students' response to training was scrutinized to measure their success. Then a composite score (Success During Training)

Table 4

Proportion of Participants in Each Treatment Group Who Responded Correctly to Each Item on the Posttest Homonym Definition Test

Homonym	Treatment (% correct)	Control (% correct)	Difference
Dance	22%	13%	+
Pen	39%	30%	+
Flower	39%	26%	+
Fish	43%	39%	+
Light	48%	35%	+
Bug	48%	35%	+
Bear	61%	39%	+
Poor	61%	39%	+
Order	74%	49%	+
Star	87%	43%	+

Table 5

Proportion of Participants in Each Treatment Group Who Responded Correctly to Each Item on the Posttest Ambiguous Sentence Definition Test

Item Type	Treatment (% correct)	Control (% correct)	Difference
(structural)	13%	13%	0
(structural)	22%	0%	+
(structural)	22%	13%	+
(structural)	35%	17%	+
(structural)	65%	26%	+
(lexical)	26%	13%	+
(lexical)	35%	26%	+
(lexical)	74%	52%	+

Table 6

Proportion of Participants in Each Treatment Group Who Responded Correctly to Each Item on the Posttest Anomaly Detection Test

Item Type	Treatment (% correct)	Control (% correct)	Difference
(micro)	48%	52%	-
(micro)	65%	57%	+
(macro)	78%	52%	+
(micro)	87%	65%	+
(macro)	61%	61%	0
(macro)	65%	57%	+
(macro)	78%	52%	+
(macro)	83%	65%	+
(macro)	87%	65%	+
(macro)	91%	70%	+

was tallied and Pearson product moment correlations were calculated. The composite score consisted of the number of items correctly and independently generated or explained by the participant across sessions. Scores on this Success During Training measure ranged from 5 to 27, with a mean of 14.79 and a standard deviation of 6.29.

In Session 1, homonyms were defined and two strategies for determining whether or not a word is a homonym were presented: 1) the form class strategy, in which students were urged to think about whether the word is both a thing and something they can do, and 2) the synonym strategy, in which students were urged to think of another word to explain their definition and see if it matches the alternate definition. The participant's task in this session was to analyze words from the word tub and classify them according to whether or not the student thought the words were homonyms or not.

Many students did not know where to begin with this task, and sat frozen in their chair, looking at the words arrayed in front of them. When this happened, the experimenter urged the student to choose a word, any word, for analysis. If the child still could not choose, the word "duck" was chosen by the experimenter. The participant was then asked what the word means. The student then typically gave either the noun or verb form of the definition, but not both. At that point, the experimenter asked the participant if the word "duck" could also be a thing (or something you could do)? Many students initially said no, thought about it, laughed, and agreed that the word was a homonym. At this point they were able to begin defining words on their own.

For the students who said no and did not realize the mistake, the experimenter ducked her head in an exaggerated manner (and/or quacked like a duck) to remind the participant of the alternate definition. This type of scaffolding was continued as long as

necessary, until the participant was able to choose a tile, read the word, define it, and successfully decide whether or not it was a homonym independently.

The number of homonyms independently identified became the first part of the Success During Training composite score. All of the students were successful in independently identifying some of the homonyms. The mean was 5.13, with a standard deviation of 2.66 (to a maximum of 22). In addition, at least two of the 23 participants trained with this program were able to read the words silently and choose homonyms from amidst the tiles, as opposed to choosing words to focus on and then deciding whether or not the word was a homonym. That said, while a few of the students professed to having heard of homonyms prior to this session, none was able to successfully define what a homonym was prior to instruction.

At the beginning of the next session students were asked if they remembered what a homonym was and if they had thought of any additional homonyms since last time. All of the students remembered what a homonym was. While few of the students reported additional homonyms, many reported that they had thought of some, but had then forgotten them, or they had made a list at home, or they had discussed it with their parents, etc.

In Session 2, participants were shown the alternate meanings of seven sentences with line drawings and then asked to build multiple meanings of sentences with Colorforms. For the most part, the students were enthusiastic about using the Colorforms for this task, especially when given specific sentences with elements to find amongst the stickers. For example, to make the sentence *The baseball bat was found by the monkey*, the child first had to search for and locate a baseball bat and monkey in the Colorforms.

This increased motivation for the task for most of the children, although there were a few who only wanted to build their own scenes with the stickers and had to be constantly reminded to remain on task.

Adding to the distraction problem was the difficulty level of the task: only six students were able to construct any of the sentences independently at first try. Instead, students would build one meaning of one form of a sentence (for example, the student would create a monkey finding a baseball bat for *The baseball bat was found by the monkey*), then the experimenter would demonstrate the alternate form to the participant (for example, a girl finding the bat as it leans against a tree from which the monkey hangs). Finally, the experimenter would give the participant another sentence of the same form to build: *The seashell was found by the penguin*. This time the student would be able to create the alternate scene independently.

The number of sentences for which the student was able to build both meanings of the sentence independently, regardless of initial scaffolding, became the second part of the Success During Training composite score. The mean score was 1.54, with a standard deviation of 1.18 (to a maximum score of 8). While students seemed to catch on to each different sentence form individually, no retention measure was built in to the training; when the students were asked at the following session if they had thought of any additional sentences with more than one meaning since the last session, not one student reported any.

Session 3 began with the experimenter reading riddles aloud to the participant. Some of the participants were delighted with this, while some were ill at ease. The experimenter made a point of explaining the features of the riddles she was reading: she

told the students that questions and answers are riddles when the same words have two different meanings, and then made a point of explaining the punch lines using words like “homonym” to explain the ambiguities.

After exploring fourteen riddles together, the participants wrote their own riddles. They brainstormed a topic and words pertaining to that topic. They looked for homonyms in the list of words. They made up questions that seemed like they would involve the primary meaning, but were really about the alternate meaning. For example, if the topic was “Harry Potter” a common word was “spell,” so an acceptable riddle might be, “Why does Harry Potter go to school? To learn how to *spell* better.” Most of the students readily thought of topics close to their hearts and were excited to make up personalized riddles. The experimenter gave them notebooks in which to record their riddles, and many of them talked excitedly about taking their notebooks home to share their riddles with Mom/Dad/brothers/sisters. All of the children made up at least two or three riddles independently that made sense, although they varied in linguistic complexity. Two children subsequently reported having written more riddles at home. The number of riddles the participants created for their topic was the third part of the Success During Training composite score. The mean number of riddles created was 2.21, with a standard deviation of 1.72. This task was constrained only by the 45-minute time limit and by students’ ability/enthusiasm. Although there was no maximum number of riddles a children could create, the most any one child wrote was seven.

In Session 4, participants read *Amelia Bedelia and the Surprise Shower* (Parish, 1979) aloud to the experimenter. Many of the children were already familiar with Amelia Bedelia, although only one child reported having previously read the specific title. There

are many instances in this book in which commonly used homonyms are misunderstood by Amelia. For example, when told Mrs. Rogers is throwing a wedding shower, Amelia gets out the hose and sprays everyone with water. As they began to read, the students understood why the experimenter had chosen this book and many spontaneously stopped to explain Amelia's difficulties understanding. For those who did not stop at ambiguity points, the experimenter interrupted and asked the participant to explain.

After reading *Amelia Bedelia and the Surprise Shower*, the participant read from *Amelia Bedelia's Family Album* (Parish, 1988). This book chronicles the professions of Amelia's various family members: for example, her cousin the boxer packs boxes for a living. The experimenter asked the participant to read the family member's name and profession, as well as Mr./Mrs. Roger's guess as to what that family member does, then covered the resolution side of these pages and asked the students to explain what Amelia will say the family member does. Having completed this, the participant brainstormed additional professions with multiple definitions and drew pictures to illustrate both. The number of resolutions correctly defined and professions independently illustrated was added to the Success During Training composite score as the fourth and final piece. The mean score was 3.08, with a standard deviation of 1.47 for the number of resolutions correctly defined (to a maximum score of 7) and 2.83, with a standard deviation of .70 for the number of professions independently illustrated. While time was the only constraint on the number of illustrations a child could create, the most any one child produced was five.

In order to assess the relationship between performance on the training tasks and verbal and literacy skills, Pearson product moment correlation coefficients were calculated and are reported in Table 7. Success During Training (SDT) correlated highly with participants' word reading ability, vocabulary, and pretest scores of reading comprehension, indicating that all of these capabilities were positively related. In addition, the SDT measure correlated strongly with some but not all of the posttest assessing metalinguistic awareness and comprehension monitoring. The SDT measure was correlated with posttests involving ambiguous sentences, riddles, and heteronyms, but not posttests involving self-corrections and anomalies. One reason may be that metalinguistic training taught students skills that impacted the first three tasks but not the latter two tasks.

Inspection of correlations among the three metalinguistic awareness tasks (see Table 7) revealed statistically significant values, with r s ranging from 0.55 to 0.70, all p s < .01. This provides evidence for multiple assessments of a single construct. In contrast, the three comprehension monitoring tasks were only weakly correlated, with r s ranging from .09 to .39, all p s > .01. This may be because different, unrelated skills were measured in these tasks. The anomaly detection task was more highly correlated with the metalinguistic awareness tasks than with the other monitoring tasks, indicating that it may assess another form of metalinguistic awareness.

Regression Analyses

Hierarchical multiple regressions were conducted to determine whether success during training explained variance in the reading comprehension posttest scores of the students who received metalinguistic training. For this analysis, participants' posttest

Table 7

Intercorrelations Between Participants' Success with Metalinguistic Training and Verbal and Literacy Abilities at Pretest

Measure (<i>n</i> = 23)	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Success during training	--												
2. Woodcock Word ID	.77**	--											
3. Woodcock Word Attack	.40	.56**	--										
4. PPVT	.73**	.62**	.16	--									

Table 7, con't

Intercorrelations Between Participants' Success with Metalinguistic Training and Verbal and Literacy Abilities at Pretest

Measure (<i>n</i> = 23)	1	2	3	4	5	6	7	8	9	10	11	12	13
5. WRMT- R Comp Pretest	.53**	.74**	.46*	.71**	--								
6. GMRT4 Pretest	.72**	.78**	.35	.67**	.69**	--							
7. Homonym Posttest	.37	.39	.20	.48*	.43*	.30	--						
8. Ambiguous Sentences Posttest	.71**	.51*	.27	.72**	.49*	.52*	.55**	--					

Table 7, con't

Intercorrelations Between Participants' Success with Metalinguistic Training and Verbal and Literacy Abilities at Pretest

Measure (<i>n</i> = 23)	1	2	3	4	5	6	7	8	9	10	11	12	13
9. Riddle Posttest	.61* *	.58**	.09	.64**	.36**	.51*	.56**	.70**	--				
10. Heteronym Posttest	.49*	.69**	.36	.60**	.69**	.51*	.56**	.48**	.39	--			
11. Self- corrections Posttest	-.02	-.06	-.11	.24	.21	-.03	-.02	.36	.18	.11	--		
12. Anomaly Detection Posttest	.30	.16	.01	.53**	.26**	.34	.35	.36	.41	.09	.09	--	

Table 7, con't

Intercorrelations Between Participants' Success with Metalinguistic Training and Verbal and Literacy Abilities at Pretest

Measure (<i>n</i> = 23)	1	2	3	4	5	6	7	8	9	10	11	12	13
13. WRMT- R Comp Posttest	.61**	.59**	.27	.67**	.74**	.69**	.54**	.44*	.41	.57**	-.11	.40	-

Note. PPVT is the Peabody Picture Vocabulary Test. WRMT-R Comp is the Woodcock Reading Mastery Test—Revised, subtest of reading comprehension. GMRT4 is the Gates-MacGinitie Test of Reading Comprehension, Level 4. WRMT-R Comp is the Passage Comprehension subtest of the Woodcock

p* < .05 (2-tailed). *p* < .01 (2-tailed).

score on the WRMT-R passage comprehension subtest was the dependent variable. The predictor variables included performance on the pretest of the Woodcock comprehension test and the Success During Training composite.

At step one, passage comprehension at pretest was entered into the equation and was significantly related to reading comprehension, $F(1, 22) = 24.90, p < .05$. The multiple correlation coefficient was .74, indicating that about 54% of the variation in posttest comprehension was accounted for by pretest ability. At step two, the Success with Training score was entered into the equation. It was also significantly related to reading comprehension, $F(1, 22) = 15.57, p < .001$. The multiple correlation coefficient with both pretest comprehension and training success in the equation jumped to .78, indicating that about 61% of the variation in the reading comprehension scores was now explained. One interpretation of this is that students' experiences during the metalinguistic awareness training sessions contributed to their reading comprehension over and above that contributed by their level of reading comprehension prior to training. Another interpretation is that students who were more metalinguistically aware prior to training responded more positively to the training.

Qualitative Analysis of Good, Average, and Poor Readers' Response to Metalinguistic Training by Reading Ability

Finally, a closer look was taken at selected students' work in order to assess what abilities or attributes were evident in student's success with the program. Participants from the metalinguistic awareness training group were categorized according to their score on the Success During Training composite: those students with scores 5-8 were considered to have achieved below average, students who scored 10-14 average, and 17-

22 above average. Natural breaks occurring in the distribution (i.e., no students received scores of 9, 15, or 16) helped to form these groupings. Two students from each group were selected randomly for greater scrutiny.

The two students with the highest scores on the Success During Training composite (22) were both boys at the older end of the age spectrum (107 and 108 months old). Although these two boys, M. and L., were from separate schools, they shared many personality traits. Each was an exceptional student, with great interest in creative endeavors (e.g., during the time of the training, M. wrote a story about a giant radioactive hot dog that was selected from many other entries at his school to be produced as a skit for a school-wide assembly). Both boys were somewhat withdrawn socially, though far from classroom outcasts. L. and M. were both extremely confident during training, eager to jump in and try out new riddles or possible homonyms. Examples of riddles written by these boys include, *Why did the Yankees play in the bakery? Because it was the only place they could find a good batter* and *Why couldn't Saint Patrick find the pot of gold? He was too green.*

Examining the pretest scores of M. and L. revealed that they scored in the highest percentile on the Peabody Picture Vocabulary Test (120 and 122, respectively). Each stated that they had prior experience with riddles, and indeed, received nearly perfect scores on all of the riddle punch line tasks (M. answered incorrectly on one of the pretest riddle items). L. also maxed out on the heteronym pronunciation and anomaly detection tasks. His pretest scores on the homonym definition and ambiguous sentence recognition tasks were at exactly the half-way point, and at posttest he showed considerable gain, improving from 5 to 7 on the homonym task and from 4 to 6 on the sentence task.

Other markers indicated that not all of the concepts were known to M. and L. prior to training. For example, during the first training session, L. asked about the difference between homonyms and homophones, and it was explained to him that homonyms are words with different meanings that are spelled and pronounced the same, whereas homophones can be spelled differently though they share the same pronunciation. At the next session, L. reported that his classroom teacher had mentioned that a pair of words she was adding to their word wall were homonyms, but he knew from our lessons that they were in fact *homophones*. This increase in metalinguistic awareness was also evident in L.'s gain scores on the homonym definition and ambiguous sentence recognition tasks: he increased from 5 to 7 homonyms correctly defined, and from 4 to 6 sentences successfully explained. As previously mentioned, there was no room for improvement in his riddle punch line task score. Likewise, there was no room for improvement in his anomaly detection ability. His score on the heteronym pronunciation task decreased, from 7 to 5, and his ability to self-correct miscues in reading could only go down: from one self-correction out of one substitution on the pretest, his posttest score fell to two self-corrections out of four substitutions. His subsequent reading comprehension gain was only one point on the Woodcock Passage test.

M., on the other hand, gained six points on the homonym definition task, two on the ambiguous sentence recognition task, and one on the riddle punch line task. He only gained one point each on the heteronym and anomaly detection tasks (and showed no gain on the self-correction task), but his reading achievement was well above the norm: from a pretest score of 489 (compared to a mean of 483) on the Woodcock Passage Pretest, he increased to 499 (the mean was 485). Thus, while the mean gain for all

participants on this standardized reading comprehension test was 2.2, with a standard deviation of 8.77, M. gained 10 points.

Two students who both earned the mean score on the Success During Training measure each gained even more than L. or M. in terms of reading achievement. K. and J., girls from different classrooms at the same school, are nine months apart in age and varied by 22 points on the vocabulary screening (with the older child achieving the lower vocabulary score). Examples of their work during training include the following riddles: *Where is Derek Jeter's home? On the baseball field* and *Why were the baseball players on strike? They couldn't hit the ball.*

Both girls scored below the mean on the pretest standardized reading comprehension tests, the word identification, and the non-word reading subtests. After receiving the metalinguistic training, however, both girls gained more than the average amount in their reading comprehension, homonym definition, ambiguous sentence recognition, and heteronym pronunciation abilities. K. and J. are both bright girls with good grades who are easily distractible and sometimes sloppy in their schoolwork. They had to be implored to focus many times during training, but when they did concentrate on the task at hand, they seemed to absorb the concepts quickly and easily. Neither of them professed any previous knowledge of the material.

Finally, A. and R., a boy and girl from different classrooms at the same school, each received a 5 on the Success During Training score, the lowest score allotted. They are of average age for the sample (105 and 102 months, respectively) and scored below the mean on the vocabulary screening (78 and 94). A. is a shy girl who often seemed a bit bewildered by the training and much more interested in the experimenter's clothing and

accessories than in the material at hand. That said, A. was especially enthusiastic about the riddle component of the training and reported sharing the riddles she had written in the third training session with her classroom teacher and parents. While she wrote riddles that were arguably not as sophisticated as some of the other students, her riddles did have an acceptable double meaning. For example, one of the riddles she devised is *Why did the photographer say “cheese”? Because he was hungry.* While cheese is not a homonym, it could be argued that in this context the photographer is not calling for someone to bring him food, and therefore “cheese” has a second meaning.

R. also wrote riddles that lacked the sophistication of some of the other students, but which still contained double meanings. For example: *What kind of dogs do sheep like? Sheep dogs!* R. is a good-natured and exuberant boy with little interest in sitting still. He scored the lowest possible score in the range of scores of the WRMT-R comprehension subtest for the experimental group at pretest and second to last on posttest; however, both A. and R. gained more than the overall average for this test.

In conclusion, children from all reading levels and language ability benefited from the training. This was noticeable not only in the gain scores of all six of these students, but in their affect and written productions as well. In addition, all of the students were enthusiastic about the training, despite a lack of enthusiasm for school in general from some of them.

Chapter 7: Discussion

Overview

This study explored the impact of a metalinguistic intervention on the reading comprehension of third-graders. One purpose was to investigate the efficacy of instruction in metalinguistic awareness for teaching reading comprehension. Based on previous research by the investigator and by Yuill (1998), it was expected that the newly-created intervention would increase participants' awareness and facility for explaining multiple-meaning words and sentences, and that this would result in superior comprehension monitoring and increased standardized reading comprehension scores. It was hypothesized that the mechanism explaining these effects would involve enhanced ability to play with the form and meaning of words and sentences, which, in turn, would make students more aware of possible alternative meanings as they read text. This would therefore increase their persistence in trying alternate meanings as well as monitoring their comprehension more closely. Closer monitoring would positively influence students' standardized reading comprehension scores.

It was also of interest to examine the effect of metalinguistic awareness instruction on readers at all instructional levels. Yuill (1998) is one of the few researchers to have conducted a similar study. However, the participants in her study were chosen specifically for having comprehension difficulties without decoding difficulties. In the present study, participants were screened for minimum decoding ability and then matched to one another based on their comprehension ability. Members of pairs were then randomly assigned to treatment group and control groups, thus assuring that both the metalinguistic awareness training group and the traditional reading comprehension group

was comprised of students with a range of reading comprehension abilities. This enhanced the external validity of the study, compared to Yuill's (1998) study.

Effects of Training in Metalinguistic Skills

The results of this study provided strong evidence that there is a causal relationship between metalinguistic ability and reading comprehension. The students who received metalinguistic training improved significantly more from pretest to posttest on the tasks of homonym definition and ambiguous sentence detection than control students did, showing that the training was effective in teaching students to think flexibly about the meanings of words and sentences. This was not the case with the riddle solving tasks, where performance did not rise above a chance level, indicating that participants were guessing at correct responses. Nevertheless, even on this task, the scores of the experimental group showed a mean gain whereas the control group showed a mean loss from pretest to posttest, revealing a tendency for performance to favor the trained group.

The strategies learned during the first two training sessions for identifying and defining homonyms and ambiguous sentences may explain what enabled the participants from the metalinguistic training group to respond more effectively on posttests than the participants from the traditional reading instruction group.

Although trained students also learned strategies for solving riddles, items on the riddle resolution posttest may have hindered the validity of that assessment. It could also be that students did not follow the instructions in this task. They were told to choose the punch line that "makes the question a riddle," and that "makes the same words have two meanings." However, many of the participants may have chosen the punch line they

found the funniest—regardless of the homonyms or structural ambiguities inherent in that punch line.

Most eight-year-olds are already familiar with the function and form of a riddle. And previous researchers (Yuill, 1998; Mahony & Mann, 1992; Hirsh-Pasek, Gleitman, & Gleitman, 1978) have considered that familiarity a benefit to assessment, since the children participating in their studies did not need much explanation and had an inherent motivation to completing the task. However, it may also be that those children who had had more previous practice with riddles better understood that the function of a riddle is to make you laugh. And what is funny to an eight-year-old is not necessarily word play of the type being assessed here. Some of the participants laughed at the alternative punch line (which did not strike the experimenter as humorous) and then chose the foil as the correct answer, despite having been instructed to choose the punch line that gives the words more than one meaning. For this reason, the riddle resolution test as an assessment of metalinguistic awareness was considered inconclusive at best. A more sensitive test, with many more items, and carefully piloted foils, is needed to determine if this type of assessment is valid and reliable.

An alternative explanation that participants receiving metalinguistic training did better on posttests could be that the training was more motivating than the training received by the traditional reading comprehension group. It is unlikely that a Hawthorne effect occurred however, because all of the training was conducted one-on-one, so all participants received individualized attention from the experimenter. In addition, the material that the control group read is a popular children's story that proved interesting to participants. Finally, the effects of the metalinguistic training were not limited to students

of any one ability level. As was shown in the qualitative review, participants from diverse reading and motivation levels gained equally from the training.

The children who received the metalinguistic training *did* on the other hand, find the training highly engaging. Their posttest performance and the qualitative examination showed that the students who received metalinguistic training were excited about words and their newfound ability to recognize word play. In this sense, the training had a motivational impact. Instead of increasing their posttest performance merely because they were proud to have been singled out by the experimenter, they were motivated by excitement for the material. In future studies, this enhanced motivation could be better controlled by training the control group with parallel material of a humorous nature. For example, the control group could read funny texts and riddles that do not rely on linguistic devices to understand the punch line (such as, *Why do fire men wear red suspenders? To keep their pants up!*)

Comprehension Monitoring Effects

On the comprehension monitoring tasks, the trained students were no more likely than control students to self-correct their word substitution errors as they read orally or to correctly pronounce each of two heteronyms contained in the same sentence. This is not surprising, as neither of these skills was taught in the training. In addition, while these tasks were created to be comprehension monitoring assessments, they are both untested by other researchers. It remains for future studies to determine the validity and reliability of these assessment tools. In contrast, students who received metalinguistic training did detect more anomalies in familiar stories than did the control students. This may be explained by the anomaly-finding tasks embedded in the *Amelia Bedelia* training session.

In that training session, the children learned to find anomalies in the text, and also to write their own multi-meaning text. This gave the participants intimate practice embedding inconsistencies into a text. However, in the case of Amelia Bedelia, the source of the inconsistencies was neither logical nor empirical. Rather it was semantic, and rested on recognizing an unintended meaning of a word. It may be that heightened sensitivity to semantic inconsistency is what facilitated performance on the anomaly detection comprehension monitoring task. In other words, recognizing that a word or phrase can have more than one meaning—and indeed, is likely to, in the Amelia Bedelia series—was enough to stimulate increased comprehension monitoring.

Of the three comprehension monitoring tasks, the anomaly detection task was also the only one that required the reader to integrate information from outside sources and other parts of the text in order to successfully complete the task. In this way, it was more closely related to text comprehension processes than were the self-correction task and the heteronym pronunciation task. The heteronym pronunciation task focused on words in isolation. While all of the sentences made sense, none were constructions produced naturally and no sentence was connected to the next. The self-correction task was a real reading task, with passages of interest to the children, but the scoring also focused on the word level—it was a measure of how often the participant recognized that he/she had substituted an incorrect word into the text and took steps to correct this error.

The divergent results of the three comprehension monitoring tasks point to the fact that much more work is needed to develop effective comprehension monitoring assessments. The most-often used method of assessing comprehension monitoring is to provide a passage with anomalous elements and ask the participants to read silently and

underline anything they find odd or incorrect. This method was pilot tested, but then rejected in favor of the more information-rich think aloud method, which uncovered some important additional evidence. For example, there were many participants who figured out the purpose of the task and reported what they found wrong at every stopping point. But there were also some children who struggled to make sense of the contradictory ideas in their think-aloud analyses.

Any time a participant made mention of the anomalous material, the item was scored as correct. It is an open question whether the students who struggled to make sense of the anomalies would have marked these items on a written test. For example, one participant said of the third anomaly in *The Three Little Pigs* (in which the big bad wolf is referred to as a bear), “Was it a wolf? Wasn’t it a wolf? It says one day a hungry *wolf* came to the straw house. Now I get it—he’s really hungry so he feels like he’s as hungry as a bear.” While there are, in fact, no indicators in the text to say that the wolf is using a humorous figure of speech, this reader did catch the anomaly, and in fact exercised constructive processing in his attempt to make sense of the text. Another benefit to this type of probe is that it was not limited by what makes sense to an eight-year-old. For example, while many of the participants debated the statement in which the wolf declared “I go to school every day,” no one commented on the impossibility of talking wolves and pigs.

Transfer to Reading Comprehension

Results were split on the question of whether training transferred to reading comprehension: students who received metalinguistic training outperformed the control students on a standardized cloze task, but not on a test requiring them to read passages

and answer multiple-choice questions. Cutting and Scarborough (2006) report that different measures of reading comprehension require different vocabulary and sentence processing abilities. For example, according to Cutting and Scarborough, cloze tasks require stronger bottom-up skills, whereas the GMRT4 relies strongly on oral language skills. Therefore, it could be that these two tests required different skills of the students.

It may also have been that the GMRT4 scores were not valid for this population of students. As reported earlier, the investigator observed that some students did not take the assessment seriously, suffered from fatigue, and did not appear to be performing their best. Although the GMRT4 is designed to be a manageable length for students this age, and the passages are about topics intended to be of interest to the population, students were still highly aware that they were taking a test. As such, they complained about the number of passages (even though the same students did not complain when asked to read *Amelia Bedelia* books of comparable length) and they ridiculed the passage topics. Students who did focus on the test and who did not attempt to find answers on other student's papers raced through the test and talked with pride about how quickly they had completed it. When asked if they had read the passages carefully, they insisted they had, but none was interested in double-checking their answers.

In contrast, in addition to being a cloze task, the WRMT-R also had the benefit of being individually administered. Although the student still read passages silently, it was not possible to look ahead in the test, nor at anyone else's answer sheet, and the investigator controlled how quickly pages were turned. This seemed to encourage much more focused attention from the participants.

The finding that students trained in metalinguistic awareness increased their scores on a standardized reading comprehension test more than their control group counterparts is consistent with the findings of Yuill (1998). Furthermore, the effects sizes calculated here ($d = .51$ for the Anomaly Detection task and $d = .73$ for the WRMT-R) are comparable with the means and standard deviations for mean improvement in months on the Neale analysis given in Yuill's published account. Yuill found that the eighteen participants in her less-skilled group who received ambiguity training improved a mean of 20.56 months, with a standard deviation of 8.07, as compared to a control group who improved 13.11 months, with a standard deviation of 7.92 ($d = .93$). The skilled group improved a mean of 13.44 months with a standard deviation of 6.73, as compared to a control group who improved 8.89 months, with a standard deviation of 10.13 ($d = .53$). Effect sizes observed in the present study are also comparable to those in other reading comprehension intervention studies. For example, Rosenshine and Meister (1994) found in their meta-analysis review of sixteen studies on reciprocal teaching that, when standardized comprehension tests were used, the median effect size was .32. Therefore, it is concluded that the magnitude of the treatment effect in the present study was strong.

More than one theory could explain the present findings. It could be that metalinguistic awareness is a new construct that directly impacts reading comprehension. Like vocabulary, background knowledge, fluency, and decoding ability, metalinguistic awareness may be an interrelated ability that is necessary but not sufficient for adequate comprehension. Another possibility is that metalinguistic awareness, as it was measured here, is an indicator of comprehension monitoring ability.

Alternatively, it could be that ambiguity training of the sort utilized in the present study produces a type of sensitivity to language that allows students to connect more completely their knowledge of spoken language to written text, and thus regulate their metacognitive control over the text. Tunmer and Bowey (1984) theorized that poor comprehenders fail to use their syntactic knowledge of spoken language when reading written text. Although they have this ability implicitly, they cannot use this knowledge to monitor their comprehension and recover from comprehension failures. Such connection forming skills are taught in this training program by cultivating students' understanding of ambiguity and by providing practice testing out multiple meanings of homonyms and ambiguous sentences, rearranging the syntax in some sentences, and reflecting on the text at hand. Unlike other reading comprehension interventions that teach strategies for text processing, the strategies taught here were embedded in the linguistic training—they were strategies for *language* processing, which indirectly boosted the children's text processing.

Cairns, Waltzman and Schlisselberg (2004) conceptualized ambiguity detection tasks as an indicator of psycholinguistic processing. Performance on tasks like the homonym definition and ambiguous sentence detection used here are thus considered symptomatic of the lexical and structural processing associated with listening comprehension. The role of listening comprehension in reading ability has not been as extensively researched as has decoding. Gough and Tunmer (1986) conceived of it as part of an interactive equation, in which $R = D \times C$, where R is reading comprehension, D is decoding, and C is listening comprehension. This means that for successful comprehension, both decoding and listening comprehension are necessary, but neither is

sufficient alone. Cairns, Waltzman and Schlisselberg (2004) therefore make the argument for the use of ambiguity detection tasks as part of the battery of tests used to diagnose reading aptitude and pre-reading difficulties at a young age.

In the current study, ambiguity detection was considered to be symptomatic of a more general metalinguistic ability, which then influenced reading through the control that was wrought over the reading process. Metalinguistic training means bringing to consciousness psycholinguistic processes that may operate subconsciously so that the reader can control and regulate them. This type of control and regulation is so similar to the thought processes that are necessary for comprehension monitoring, it may have been that students trained in ambiguity detection were then better able to monitor and repair their comprehension. This was evident in the significant gains on their homonym, anomaly detection, and reading comprehension post-tests. This reprocessing is hypothesized to have been what occurred as well in the miscue self-correction task, however the procedural and scoring issues as discussed above kept the task from revealing useful information.

These results could have far-reaching implications. Metalinguistic awareness, conceptualized as phonological awareness, is well-known to be an indicator of future reading success (see studies such as Bryant and Bradley, 1985), but this study represents a new direction in metalinguistic awareness. Ambiguity detection is metalinguistic in that students must consciously wield control over their mental processes in order to recognize the double meaning of words and sentences, and reprocess that meaning. That skill, however, does not appear to correlate with decoding skill, as phonological awareness does, but rather with comprehension. More research is necessary to determine

if this is because ambiguity training increases students' processing ability and thus their listening comprehension, or if merely by strengthening students' understanding of the reading process, they were better able to control their comprehension monitoring, and thus recognize the need for reprocessing.

Limitations and Future Research

A larger scale study into this type of metalinguistic instruction should include a diverse population of students. The two schools included here were both mid-sized elementary schools in a large urban district. The population of participants sampled here was somewhat narrow. Only third-graders without IEPs were included in the present study. In Yuill's (1998) study, the children were slightly younger; they were 7 years-old, on average, as opposed to 8 years-old here. The present study looked at a broader population, which allows for more generalization. However, an even greater range of participant ages and abilities needs to be included to strengthen external validity.

Another weakness is the untried nature of the assessment procedures both in the area of metalinguistic awareness and comprehension monitoring. More research is needed into the best practices for assessing these constructs. As mentioned above, the think-aloud protocol was particularly useful in this investigation for the information-rich nature of the presentation. It was not, however, the most efficient measure, in that it took a fair amount of time for students to read the stories aloud and to speak their mind. At that point, the tapes of student talk still had to be transcribed, classification schemes developed, and the content analyzed, then scored by two independent raters.

In addition, the think aloud measure in general has been criticized for altering the activity it purports to assess by halting the activity and imposing an additional cognitive

burden (Schooler, Ohlsson, & Brooks, 1993). Along these lines, Ericsson and Simon (1980) have argued that the perfect subject is one who is capable of an almost trance-like unconscious verbalization of thoughts while remaining completely focused on the task at hand. The participants in the current study did not achieve this level of focus: many of the participants recognized early in their readings that there would be anomalies to find in the texts and happily stopped to point them out. Others had to be interrupted by the experimenter and asked to stop reading in order to think aloud. Despite this, the think aloud task comes the closest to assessing authentic text reading from among those available. An alternative explanation is that a heretofore-undefined construct mediated the correlation between the rise in scores on the metalinguistic tasks and the standardized reading scores, rather than comprehension monitoring. For example, Yuill (1998) suggests that there is more work to be done exploring the relationship between working memory and poor comprehension. Further research is needed into the nature of this relationship.

Finally, the divergent results of the three comprehension monitoring tasks point to the fact that there is much more work to be done developing effective comprehension monitoring assessments. Ruffman (1996) theorized that the type of textual problem employed affects the extent to which a task truly measures comprehension monitoring. For example, he points to the following empirical inconsistency from Markman and Gorin's work (1981, p.321): "They [koalas] sleep on the tree tops where cool, soft grass grows." Ruffman argues that if a reader were to come across this sentence in a text, he would probably note the inconsistency and regard it as an error of the author, but argues that it would not cause comprehension failure. Logical inconsistencies, on the other hand,

are incompatible within the text and therefore must cause comprehension failure if not resolved. He offers this example, again from Markman and Gorin (p. 321), “[Koalas] will sleep only high up on the tops of trees... They sleep on the ground in the cool, soft grass.” In this example, it is unclear which proposition is correct, and in the absence of background knowledge, the reader is left unable to comprehend. Ruffman’s conclusion is that, although there may be situations in which empirical inconsistencies threaten a reader’s comprehension, logical inconsistencies are a better means of assessing comprehension monitoring.

Since the micro and macro prompts in this study were analogous to Markman and Gorin’s logical and empirical inconsistencies, a fruitful future analysis might be to compare the think aloud responses of the participants in this study on the micro prompts to those on the macro prompts. As mentioned earlier, the macro prompts were harder to score in this study; interrater reliability diverged mostly on the macro items, when students were more likely to mention the item, but made irrelevant comments. In some cases, the children clearly brought personal experience to the task, but were not necessarily aided by that personal experience. For example, in responding to the prompt, “Bears like to eat jellybeans,” one girl responded quickly, “I think Momma Bear’s been hiding jellybeans under her pillow even though she’s not supposed to be eating after dinner.”

Conclusions and Implications

There are many subprocesses that contribute to successful reading. Gough, Hoover, and Peterson (1996) have listed some of the things readers must do. For example, readers must control their eye movements, recognize words, disambiguate the

words, fit the words together into appropriate syntactic relations, determine the sentence meaning, make inferences about the text meaning, relate that sentence to the previous text, fit the text to a mental discourse, and finally decide what to do with the information. All of these things happen in a manner of seconds. If just one of these processes fails, comprehension fails. Training in metalinguistic awareness is one way to learn to control and regulate all of those processes, to ensure successful comprehension.

The results of the current study combine with Yuill's (1998) findings to suggest that reading comprehension can be taught (or at least reinforced) through a short, simple training, based on language games and riddle manipulation. This study provides convergent evidence for the idea that being flexible with language results in better reading comprehension. Unlike Yuill (1998) however, whose participants were chosen for their relative lag in comprehension abilities compared to their decoding skill, the participants in this study ranged in abilities and thus more closely represented the diversity found in a real classroom.

The findings of this study could therefore have a significant impact on the way in which educators design comprehension instruction: in addition to teaching strategies, perhaps it is necessary to teach children how to think about language.

Appendix A

Written Consent Form

Dear Parent:

My name is Marcy Zipke. I am a graduate student in the Educational Psychology Ph.D. program at The Graduate Center, CUNY. To fulfill the requirements of my dissertation, I will be conducting a research study at your child's school this fall to determine the most effective methods of teaching reading comprehension. I am writing to invite your child to participate.

I will be tutoring students during school hours. First, I will give students some reading and writing activities to do in order to see how far they have progressed in learning to read. Sixty students who can read the words on the page, but who have difficulty comprehending what they read will be randomly chosen to receive tutoring for a period of about four weeks. This tutoring will take place during school hours, at times the classroom teacher deems least objectionable to removing the children from the classroom. Half of the students chosen to participate will receive six tutoring sessions, focusing primarily on language. The other half will receive two small-group sessions, focusing on the larger aspects of a story, such as plot or setting. Each session will last about 45 minutes.

All of the sessions will be audio recorded, in order to better study the impact of the instruction. There is no cost for this tutoring, and there are no known risks for the children who participate. On the contrary, all of the children who participate should

benefit from the instruction by improving their ability to read. In addition, I will be happy to share with you any information I uncover about your child's strengths and weaknesses in reading. This information will be kept in a locked file in my office, to which only my advisor and I have keys. No one else will have access to that information without your written consent. The identities of all participants in the study will be kept confidential. Participation in this study is voluntary. Participants are free to withdraw from the study at any time without consequence; parents should contact me if they wish to discontinue a child's tutoring.

If you have any questions about the project, please feel free to call me at 718-832-3147 or email me at mzipke@gc.cuny.edu or to contact my advisor, Dr. Linnea Ehri, at 212-817-8294, lehri@gc.cuny.edu. If you have any questions concerning your child's rights as a participant in this project, you may contact Kay Powell, IRB Administrator, CUNY Graduate Center, 212-817-7525, kpowell@gc.cuny.edu.

Thank you for considering this request. Please print your child's name and check the box below if you agree to let him/her participate.

I grant permission for _____ to participate:
(Child's name)

I give permission for my child to be audio taped. Circle one: Yes No

(Signature of parent/guardian)

(Date)

Best regards,

Marcy Zipke
CUNY Graduate Center

Appendix B

Script for Obtaining Oral Consent from Participants

My name is Marcy Zipke. I am a student, just like you. I go to CUNY. Have you heard of that school? I am doing a research project. I am trying to figure out the best way to help children learn to read better. To do this, I am going to read with some of the children in your class, but not all of them, for the next few weeks. I am ready to begin working with you, but first I need to tell you what we will be doing, and you need to tell me that you want to work with me.

[FOR EXPERIMENTAL GROUP CHILDREN:]

My project is all about helping children read better. Today, I will teach you some new words. Next time we meet, I will teach you some new sentences. Then, we will meet a couple more times and we will make up riddles and read books. I will also ask you to do some reading and writing.

[FOR CONTROL GROUP CHILDREN:]

My project is all about helping children read better. Today, we will be reading books together. We will talk about what is in the book. We will meet a few more times and I will ask you to do some more reading and writing.

It is up to you whether or not you want to work with me. You can stop working with me any time you decide, and that will be okay.

Would you like to work with me?

Oral consent granted by student? Yes No

If child denies consent, express thanks and dismiss child from study.

Appendix C

Homonym definition test: Metalinguistic awareness

I am going to read you some words. I want you to tell me what it means. If it can mean more than one thing, tell me all the meanings you can think of. Ready?

Pretest Homonyms:

1. Bow
2. Bank
3. Feet
4. Mouth
5. Sink
6. Treat
7. Check
8. Can
9. Shed
10. Run

Posttest Homonyms:

1. Star
2. Poor
3. Light
4. Bear
5. Pen
6. Flower
7. Fish
8. Bug
9. Dance
10. Order

Appendix D

Ambiguous sentence comprehension task: Metalinguistic awareness

Have you ever thought about the fact that sometimes sentences have more than one meaning? Here's an example:

The chicken was ready to eat.

One meaning of the sentence is, The chicken is hungry and ready to eat his dinner. [Show child picture]



Another meaning is, The chicken is cooked and ready for other people to eat. [Show child picture]



Here is another example: "They talked about the problem with the teacher," can mean:

- They talked to the teacher about a problem.
- or
- They talked to someone else. They talked about a problem they were having. The problem had to do with the teacher.

I'm going to read you a sentence and I want you to tell me all the possible meanings of the sentence. Start by telling me the first meaning that comes to mind, then I'll ask you if the sentence can possibly mean anything else.

Pretest Sentences:

1. The man's nails were very sharp.
2. The elephant was ready to lift.
3. The cold made Betty feel terrible.
4. The boy watched the little fish and turtles.
5. The glasses fell on the floor and broke.
6. The sheriff caught the man with the gun.
7. The boy picked up the bow.
8. Flying kites can be exciting.

Posttest Sentences:

1. The woman saw the broken cups and dishes.
2. The children saw a bat lying by the fence.
3. Bouncing balls can make people laugh.
4. The nurse looked over the chart.
5. The fat soldier's wife was standing by the window.
6. The children showed the man the straw.
7. The girl tickled the baby with the stuffed animal.
8. The man held the pipe.

Appendix E

Riddle resolution: Metalinguistic awareness

Instructions (to be read aloud to participants):

A **riddle** is a question that turns into a joke. It starts with a puzzling question and ends with an answer that surprises you and usually makes you laugh. The answer to the riddle is called the **punch line**. The question and answer make a riddle when the same words have two different meanings.

Here are some questions. I am going to tell you two possible punch lines after each question. One punch line makes a riddle and the other does not. Your job is to decide which punch line makes the question a riddle.

Pretest Riddles (correct answers labeled “a”):

1. Why should you never swim on a full stomach?
 - a. It’s easier to swim in water
 - b. You’ll get sick.

2. Why didn’t anyone take the bus to school?
 - a. It wouldn’t fit through the door.
 - b. It wasn’t cool.

3. What kind of animal can jump higher than the Empire State Building?
 - a. Any animal—the Empire State Building can’t jump!
 - b. None.

4. How many sheep does it take to make a sweater?
 - a. I didn’t even know they could knit!
 - b. Ten.

5. Why did the skeleton go to the movies by himself?
 - a. He had no body to go with him.
 - b. He was lonely.

Posttest Riddles (correct answers labeled “a”):

1. Do you sleep on your stomach?
 - a. No, on a bed
 - b. Yes.

2. Why did Frog eat a lamp?
 - a. He wanted to eat a light snack
 - b. He was hungry.

3. Why was the policeman in bed?
 - a. He was an undercover cop.
 - b. He was tired.

4. Why is a school yard larger at recess than at any other time?
 - a. At recess there are more feet in it.
 - b. It isn't.

5. What kind of house weighs the least?
 - a. A lighthouse.
 - b. An apartment.

Appendix F

Heteronym pronunciation: Comprehension monitoring

Instructions: Please read these sentences out loud to me.

Pretest Heteronyms:

I can't read that book next because I've already read it.

She was content to read about the content of the book.

The police had to permit the party because the people had a permit.

He crooked his neck to see the crooked tree.

Because of his wound, he had bandages wound all around him.

In order to object, Sam had to move the object.

He took the lead in picking up the lead pipes.

Sam waited until after dinner to present her with the present.

I lowered my bow and arrow and took a bow.

When the fire was close, the police had to close the road.

Posttest Heteronyms:

How does a student know which deer are does?

I found a live bear where I live.

I couldn't wind up the hose because of the strong wind.

The tear in the fabric caused Mom's eyes to tear up.

The maid will separate the laundry into separate piles.

Giving a good excuse sometimes helps excuse you from punishment.

I subject my friends to pain when I discuss the subject of my operation.

The perfect teacher is one who helps you perfect your skills.

We would shop at his store if he could produce the fresh produce.

She dove into the water like a dove in the air.

Appendix G

Miscue self-correction: Comprehension monitoring, task passages

Instructions: Please read these stories out loud to me. I will tell you when to stop reading.

Pretest Passages (from O'Donnell & Wood, 1989):

Passage 1:

Bob had a new red bike. He liked to ride it. One day he stopped to show it to his friend, Ann. She liked his bike. "I wish I had a bike," she said. "Then we could ride together."

Passage 2:

I met Tim on the way to school. He had a box. There were three frogs inside. We stopped to play with them. I picked one up. So did Tim. We did not see the other one jump out of the box and hop into my school bag. I grabbed my bag. We ran all the way to school. When I got to school, the frog jumped out of my bag. Everyone laughed.

Passage 3:

It was a cold rainy day. Everything went wrong. Mary spilled milk on her new pants. Her mother yelled at her. Her little brother pulled the eyes out of her teddy bear. Her sister called her a stupid brat. Mary was so mad that she decided to run away. She put some cookies in a bag and went outside. She walked in the rain for a long time. She tried to eat

her cookies but they got all wet. Finally Mary went home. Her mother gave her a big hug. She began to feel better.

Passage 4:

In the middle of a large field there was a hole that ran down to a hidden room underground. A family of woodchucks had dug it out early in the fall. Deep in the safe dark hole were five new babies. They were all snuggled together. They were sleeping on a soft bed of dried hay.

When they were first born, all the babies wanted to do was sleep. Their eyes were still closed, and they didn't move around at all. But after a few hours their eyes opened. With their small black noses they sniffed all around. They were all hungry.

Mother Woodchuck gathered her babies close to her. They drank her warm milk.

Passage 5:

Elephants live in the wild in some parts of the world. There used to be many wild elephants. But people are moving onto land where elephants have always lived. People are building homes and towns. The elephants don't have enough land to live on or food to eat. Many of them are dying.

Another reason wild elephants are disappearing is that hunters are killing them for their ivory tusks. They sell the tusks to people who make them into expensive combs and jewelry. Things that are made from ivory are sold all over the world.

Some people are trying to save the elephants. They are making special parks for elephants to live in. They are also passing laws to keep people from killing too many elephants. They are asking everyone not to buy things that are made from ivory.

Passage 6:

Once upon a time a big lion was asleep under a tree. A little mouse came by and started playing with his tail. The lion woke up and grabbed the mouse with his paw.

“How dare you play with my tail?” he roared. “I am going to eat you up.”

“Oh please, Lion, don’t eat me,” begged the mouse. “If you will let me go, I will help you some day.” The lion laughed. “How could a little mouse help me? That is such a funny idea that I will let you go.” And the mouse ran away.

A few days later, some hunters left a net under a tree to catch the lion. The lion got caught in the net. He could not get free. He roared and roared. The mouse heard him roar and ran to see what was wrong. “Lie still,” said the mouse, “I will help you.” He chewed on the net with his sharp little teeth. He chewed and chewed. He chewed a big hole in the net and the lion got free.

“You see?” said the little mouse. “I did help you.”

Passage 7:

In 1946, the game of baseball was over 100 years old. But there were no black players on any major league teams. Branch Rickey, manager of the Brooklyn Dodgers, wanted to change this. He started looking for a black player to lead the way. The man had

to be special. He had to be a very good player. He had to be brave. Many fans would call him names and boo him. Other players might refuse to play with him/

Rickey finally found his man. Jackie Robinson had been a sports star in high school and college. He was playing on an all-black team when he agreed to play for the Dodgers.

Robinson's problems began right away. Some teams didn't want to play with the Dodgers if Robinson played. But the head of the league said they had to. Robinson couldn't believe how much some of the fans and other players hated him. They called him names. He wanted to fight, but he promised not to. He fought back by staying cool and playing hard.

In his first game, he scored the winning run. He played well all season, but it was very tough. A turning point came when a runner from another team smashed into Robinson's leg with his sharp spiked shoes—a dirty play. Robinson's whole team ran onto the field, ready to fight. They showed they were behind him. He was part of the team.

Robinson played for many years. He was so good that when he retired, he was voted into the Baseball Hall of Fame. He showed everyone he was a hero.

Posttest Passages (source unknown):

Passage 1:

Patty has a green bird. Mikey can say his name. He likes to yell and make lots of noise. He can be a pest.

Fran has a yellow cat. Whiskers has no tail. She jumps and hops. When she runs, it is funny to watch.

Shannon has a black dog. Lady's tail was clipped. It is bad when she chases cars and trucks. She might get hurt one day.

Pets can make our lives more fun. We must take good care of them. They need us even when they are being bad.

Passage 2:

My cat is not like my dog. She is so different. My cat lives indoors. My dog lives outside.

My cat likes to sit on laps and purr. My dog likes to run outside and fetch sticks.

My cat can do whatever she likes. It does not matter what we think. My cat does not do tricks.

My dog does do tricks and loves to please. When we are happy, he wags his tail. He can shake hands, roll over, and fetch sticks.

Passage 3:

In a castle, far away, lives the cat with golden fur. Her mistress is a princess. Her best friend is a pony. The pony belongs to the king.

She loves to nestle in the lap of the princess. She loves to get cozy and purr. The princess pets her ever so gently. She pets the cat with the golden fur.

The cat with the golden fur is one of a kind. There are none like her in the whole world. In fact, there will never be another like her. She is very, very special.

The cat is a royal secret. Only the king, the queen, and the princess know about her. They all love the cat with the golden fur. They keep her safe and sound.

Passage 4:

Do you catch colds every winter? If you do, then there are things you can do to help prevent them; things certainly worth the try.

During the cold season it is wise to wash your hands several times a day. The cold germs on your hands will then be washed down the sink. Better there than in you.

You need to get out of the bad habit of rubbing your eyes. Many doctors think that is one of the main ways you catch a cold. The cold virus enters your body through the eyes. Soon, you become sick with a cold.

You can eat fresh fruit and take Vitamin C to prevent colds. Stay in shape with exercise and get plenty of sleep each night. If you are strong, you may be able to fight off infection.

By doing these things, you may avoid a cold next year. In fact, you may never catch another cold in your life. Now wouldn't that be nice?

Passage 5:

Bob is my best friend, but Bob brags too much. If I tell Bob about my best fish, you can bet he got one better. He always has to do you one better.

Should I have some fun and tell him about a fish too big for there ever to be such a thing? Should I make up a fish tale that would be hard for him to better?

I think I should. I think it would serve him right. Maybe that would make Bragging Bob work on his bad ways. I could see just how far he's be willing to go to best my fish story.

Yes, that's what I'll do. It may take me a while to dream it up, but when I do... My fish tale will be a whopper. Or should I call it a whaler?

Passage 6:

American sailors were enjoying a dreamy Sunday morning in Hawaii. They could sleep late—take things a bit easy. Life in the navy was just a bit nicer on Sundays. Things were relaxed; perhaps a bit too relaxed. This was December 7, 1941; a day President Franklin Roosevelt would call “a day of infamy.”

Japan and the United States were on very unfriendly terms. There were formal discussions going on in Washington, D.C. to see if a way to peace could be found. Little did the Americans know that a Japanese attack armada was preparing to destroy the American Pacific fleet at Pearl Harbor.

Japanese fighters and bombers attacked with the first light of morning. The Americans at Pearl Harbor were caught by complete surprise. Many ships were sunk or crippled. Planes were destroyed before they could take off. Thousands of soldiers died.

By the end of the day, it was obvious that the United States had been caught off guard. The Pacific fleet had been devastated. The United States declared war on Japan and vowed never to forget the lessons of Pearl Harbor.

Appendix H

Error detection: Comprehension monitoring

Instructions: Here is another story for you to read out loud to me. This time I want you to stop every time you come to one of these stars in the story (show child the ***) and tell me what you are thinking. Just tell me whatever you are thinking about when you get there. You might be thinking about what is going to happen next, or about something similar that happened to you, or you might be confused about what you just read... whatever you are thinking, you tell me. For example, if I were to think aloud while I read the story you just read, I would read, "Bob had a new red bike," then I would stop and I would think out loud about MY new bike, because I just got a new bike too, and that is what *I* would be thinking about. Do you understand?

Pretest Story:

The Three Little Pigs

Once upon a time, there were three little pigs who lived near a forest. They were all brothers. They liked to eat and play all day long. Winter was coming soon, when the weather would be very hot.*** They decided to build themselves some houses.

The first little pig built a house out of straw. The second little pig built his house out of sticks. The third little pig built his house out of bricks. The fourth little pig built his house out of mud. ***

One day a hungry wolf came to the straw house. He knocked on the door.

"Who is it?" asked the little pig.

“I’m a big bad wolf!” said the hungry bear. ***

“Go away!” said the little pig.

“Little pig, little pig, let me in,” said the wolf.

“Not by the hair of my chinny chin chin,” said the pig.

“Then I’ll huff and I’ll puff and I’ll blow your house in.”

The wolf huffed and puffed and the straw house melted. *** The first little pig ran to the house made out of sticks.

“What’s wrong?” asked the second pig.

“There’s a tornado coming!” said the first pig. ***

Just then there was a knock.

“Who is it?” asked the two little pigs.

“I’m a big bad wolf!”

“Go away,” said Momma Pig. ***

“Little pigs, little pigs, let me in.”

“Not by the hair on our chinny chin chins.”

“Then I’ll huff and I’ll puff and I’ll blow your house in.”

The wolf huffed and puffed and the house made of sticks blew apart. The little pigs ran to the house made out of bricks.

“What’s wrong?” asked the third pig.

“There’s a wolf coming!” said the first pig.

“Pigs eat wolves for breakfast,” said his brother. ***

Just then there was a knock.

“Who is it?” asked the three little pigs.

“I’m a big bad wolf! I go to school every day.” ***

“Go away,” said the pigs.

“Little pigs, little pigs, let me in.”

“Not by the hair on our chinny chin chins.”

“Then I’ll huff and I’ll puff and I’ll blow your house in.”

The wolf huffed and he puffed and he blew. But nothing happened. The wolf gave up and went away.

The ten little pigs came out of their house made of brick. They danced and sang, “Who’s afraid of the big bad wolf?” ***

They lived happily ever after in the house made of mud. ***

Posttest Story:

The Three Bears

Once upon a time, there were three little bears. There was a mama bear, a papa bear, and a baby bear. These bears lived in a house in the woods.

One day Mama Bear was making porridge for breakfast, but it was too hot. The bears decided to go for a walk in the woods while the porridge cooled off.

While they were out, a little girl with long golden hair came to the door of the house and knocked. Her name was Goldilocks. She knocked and knocked, but no one answered. Cinderella was lost and tired, and very hungry. ***

Finally, she peeked inside. “Hello?” she asked, but no one answered. Goldilocks went into the dining room where she found the three bowls of porridge on the table. She tasted Papa Bear’s porridge with a pencil, but it was too hot. *** She tasted Mama Bear’s

porridge with a spoon, but it was too cold. She tasted Baby Bear's and it was just right. She ate it all up.

Then she decided to sit down. First she sat in Papa Bear's chair, but it was too hard. Then she sat in Mama Bear's chair, but it was green. *** Finally, she sat in Baby Bear's chair and it was just right. She started to read a book, but then Baby Bear's chair broke into a million pieces.

Goldilocks was very tired now. She went upstairs to look for a bed to lie down in. First she tried Papa Bear's bed, but it was too hard. Then she lay in Mama Bear's bed, but it was too soft. Finally, she lay in Uncle Bear's bed, and it was just right. *** She fell fast asleep.

A little while later, the three bears came home from their walk in the field. *** They went into the dining room and sat down to eat their breakfast.

"Somebody's been eating my porridge," said Papa Bear.

"Somebody's been eating my porridge," said Mama Bear.

"Somebody's been eating my spaghetti," said Baby Bear, "and they ate it all up!"

The Bear family went into the living room to talk about what to do.

"Somebody's been sitting in my chair," said Papa Bear.

"Somebody's been sitting in my chair," said Mama Bear.

"Somebody's been sitting in my chair," said Baby Bear, "and they broke it into pieces!"

By now, the Bear family was upset. When upset, most bears watch TV.*** The family went upstairs to check the rest of the house.

“Somebody’s been sleeping in my bed,” said Papa Bear.

“Bears like to eat jellybeans,” said Mama Bear. ***

“Somebody’s been sleeping in my bed,” said Baby Bear, “and she’s still there!”

“MEOW!” said all the bears. ***

Goldilocks woke up and ran downstairs, with the bears chasing her. Goldilocks explained to the bears that she was lost and hungry and very, very sorry. She helped Papa Bear to fix the chair, she helped Mama Bear to make more porridge, and she taught Baby Bear how to make his bed. Then Goldilocks washed the dishes and the Bear family gave her directions from their home near the forest in Brooklyn. ***

They all lived happily ever after.

Appendix I

EXPERIMENTAL GROUP SCRIPT and MATERIALS (All materials to be used in training—words, sentences, riddles, etc.—built into script)

Session One

My name is Marcy. Remember the last time we met I told you that I'm trying to figure out what is the best way to teach children how to read? Well, I'm going to be seeing some of the children in your class, but not all, for the next few weeks, to see if I can help them understand what they read a little bit better.

Here is what is going to happen: You and I are going to meet each week and we're going to play games with words. I expect you to try your best and think really hard about everything we do. Can you do that? ... Good.

Let me tell you about something that happens to me sometimes when I read. Sometimes, when I read, I can read a whole sentence and even though I'm able to *read* every word, I get to the end, and I still don't know what the sentence meant. Does that ever happen to you?

Sometimes that happens because I wasn't paying enough attention while I was reading. But sometimes that happens because the words are tricky. Let's talk about what you can do when you read a sentence just fine, but you don't understand what it means. What do *you* think is the first thing you should do? [Scaffold responses until children say re-read sentence.]

Excellent! Sometimes, you have to look at those words again, don't you? Even though you may know what a word means, it may have fooled you. Well, here's a big word for you... Have you ever heard the word "homonym" before? [If any child says yes, let her/him explain what s/he thinks a homonym is.] A homonym is a word that has more than one meaning. Like "watch." You know what the word "watch" means, right? What does watch mean? [Children will explain either verb or noun definition, but probably not both.] Yes! That's right. "Watch" means [one definition]. The tricky thing about "watch" though, is that it can also mean [alternative definition].

Can you think of any words that have more than one meaning? [If the children need help, encourage them to look around the room for words.]

Good job! Now I'm going to give you a few examples and I want you to tell me if the definitions I give make the word a homonym or not:

1) First we'll try WATCH. I can *watch* a movie or I can *watch* you. Are those definitions different? Noooo...[get children to explain that in both of those you open your eyes and

try to see something] But what about, I can *watch* a movie or I can read the time on my *watch*. Are those different?

[Introduce form class strategy—ask the children to think about if it’s a thing, and if it is something they can do.]

2) Good. Here’s another one like that: BEAR. I could say one is a grizzly bear and one is a polar bear, but are those really different? Both are still big furry animals... How about trying the strategy we just used for watch? We know bear can be a thing, but is it also something you can do? [I can’t bear to go out in the rain!] Make sure you don’t convince it with this word [show BARE] because that kind of bear is spelled differently...

3) Let’s try a different kind of homonym. How about BALL? The ball could be a baseball or it could be a soccer ball. Are those different? [Explain that while they do look a little different, they are still both round and filled with air.] How about a baseball and the kind of ball that Cinderella went to?

[Introduce synonym strategy—what’s another word for a baseball? Toy. What’s another word for Cinderella’s ball. A dance.]

4) Here’s the last one: COLD. I could say that it’s cold outside, or that the water in my water bottle is cold. But those aren’t so different, are they? Tell me another word for cold, like cold outside. [Wait/scaffold for chilly or freezing.] Is that another word you could use to describe my water? [Yes.] Good!

Now we’re going to look at the words in this tub. Your job is going to be to find the words that have more than one meaning. Use the strategies we talked about.

Let’s see if each of you can find all of the words with more than one meaning. [Put homonyms off to the side where children can reference them.]

AMBIGUOUS WORDS IN THE TUB:

DUCK
FISH
LIKES
FLIES
WORK
BALL
RIDE
WATCH
COLD
LIVES
PLAY

READ
HOT
GOOD
BOOK
HOUSE
SCHOOL
FLOWER
HELP
BIKE
TRUCK
COLOR

You did a fantastic job finding all of those homonyms! I'm going to see you again on _____. Try to think of some more homonyms that you can tell me on _____.

Session Two

Last time we met, we talked about homonyms. Do you remember what a homonym is? Can you tell me what a homonym is? [Respond to answers appropriately.] A homonym is a word that has more than one meaning. Did you think of any new words with more than one meaning since we last met? Good. Well, today we're going to look at some sentences with more than one meaning.

I'm going to tell you some tricky sentences, and then show you some pictures. [Tell children the sentence first, then show the illustrations and let them take turns explaining how the pictures illustrate the sentence.]

SENTENCE LIST (with example illustration):

1) The chicken is ready to eat.

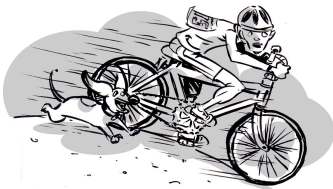


Illus.#1:



Illus.#2:

[In the first picture, the chicken is hungry and wearing a bib, so he must be ready to eat his own dinner. In the second, the chicken is cooked and on a plate, so it must be ready for someone else to eat.]



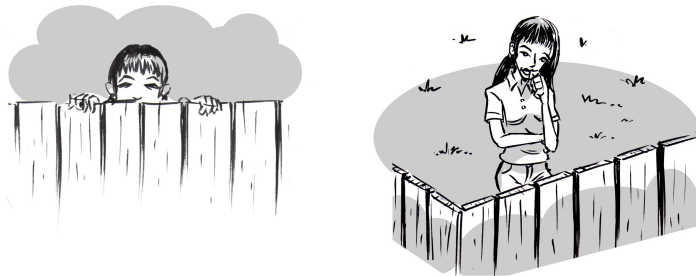
2) The dog chased the man on a bike.

[In the first picture, the man is riding the bike and the dog is chasing him. In the second, the dog is doing the chasing *on a bike*.]



3) Drink your juice after your bath.

[In #1 he has just gotten out of the bath, and is drinking his juice. In #2 he is drinking his bath, then he will drink his juice.]



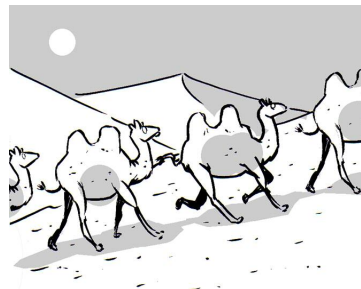
4) The woman looked over the fence.

[In #1 the woman is peering over the top of the fence. In #2 the woman is looking over the fence like a teacher looks over homework.]



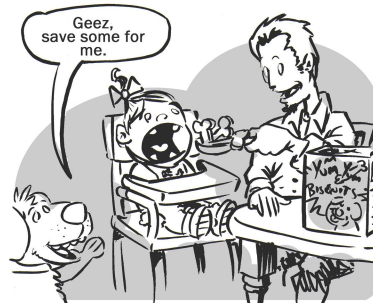
5) The doctor helped the snake bite victims.

[#1 shows a doctor helping someone who has been bitten by a snake. #2 shows the doctor biting victims for the snake.]



6) The two(-) humped African camels were walking across the desert.

[#1 shows two camels with 1 hump each. #2 shows multiple camels with two humps.]



7) He fed her dog biscuits.

[#1 shows a man, a girl, and her dog—the man is giving her dog the biscuits. #2 shows the same characters but the “her” refers to the girl and the man is giving her *dog biscuits*.]

Great job with those sentences! Now we’re going to play with these stickers. Have you ever seen these before? They’re just like stickers, except that you can peel them back up again and stick them back down. Let’s practice...

Here’s what we’re going to do with the stickers. I’m going to tell you a sentence, and I want you to first think about how many meanings you think the sentence might have. Then we’ll try to make the meaning with the stickers.

Let’s start with, “The girl played soccer.” How many meanings do you think that sentence has? [1] Good! Let’s make a picture of that sentence with the stickers.

Now let’s do a harder sentence. How about, “The ball was found by the kitten.” How many meanings can you think of for “the ball was found by the kitten?” [Show sentence card and build with Colorforms as follows to demonstrate:



We’re going to write down the meaning of each sentence on this chart. [Show students the chart with a column for WHO, a column for WHAT and a column that says WHOM.]

On this line, we’re going to write down *who* is doing the action in the sentence. On the next line, we’ll write down *what* they are doing. And on the last line, we write down *whom* they do it to. I think it will make more sense if we do it with a real sentence, so let’s try the sentence we just made.

[Prompt child with these questions:]

Look at your picture and tell me who is doing the action.

What is it that ____ is doing?

And what did she find?

Great job!

WHO	WHAT	to WHOM
Kitten	Found	The ball
Girl	Found	The ball (by the kitten)

Here is your next sentence [read sentence from list below]. Let's make a picture of that sentence with the stickers... Good work! Let's put it on your 3W chart: Who did what to whom in the sentence?

Now, can you think of another meaning for that sentence? Try putting the stickers down in a new way to make the sentence again. [Pause to let the child reassemble the pieces of the sentence. If he/she can't do it, put the stickers down for him/her.] Okay! What does it look like now? Let's put the sentence on your 3W chart again. Who did what to whom now in the sentence? Good. Let's try another one like that.

AMBIGUOUS SENTENCES FOR COLORFORM EXERCISE:

- The baseball bat was found by the monkey.
- The seashell was found by the penguin.

- A little boy and girl went for a walk.
- Purple boys and girls were playing baseball.

- The wrench might have been buried with the shovel.
- Maisie talked to Bob the Builder with Little Bear.

- Maisie likes visiting relatives.
- Little Bear has to be careful to avoid biting snakes.

You did a great job building those sentences! We'll do some more next time I see you. I'd also like you to try to think of one or two sentences with more than one meaning that you can bring to me for our next meeting.

Session Three

Today we're going to play with riddles. Do you know what riddles are? A riddle is a question that turns into a joke. It starts with a puzzling question and ends with an answer that surprises you and usually makes you laugh. The answer to the riddle is called the punch line. The question and answer make a riddle when the same words have two different meanings. Do you have a favorite riddle? I have some favorites... Let's see if we can figure out what they mean together. We can use our 3W chart to write down each of the meanings of the riddles.

[Read some riddles with children, try to gauge whether or not they understand the alternate meanings on their own.]

Riddles

1) Why do spiders like baseball? They're good at catching flies.

[Explanation: "Flies" is a homonym. It can mean balls hit to the outfield in baseball, but that is not the same kind of "flies" that spiders eat.]

2) What has a mouth but can't talk? A river.

[Explanation: "Mouth" can be a body part or the opening of a river, where it meets the sea.]

3) Why are fish so smart? Because they swim in schools.

[Explanation: "Schools" can mean a place for learning or a group of fish.]

4) Why did the witch go to night school? She wanted to learn how to spell better.

[Explanation: "Spell" that can mean the letters in a word or the magic a witch uses.]

5) Do you have any fans in your house? No, everybody hates me.

[Explanation: "Fans" can mean blades that move the air or someone who is crazy about a person or thing.]

6) Why can't cheetahs hide very well? Because they're always spotted.

[Explanation: "Spotted" can mean seen or with polka dots.]

7) How do you stop a skunk from smelling? Hold its nose.

[Explanation: "Smelling" can mean give off an odor or sniff at something.]

8) Where can you see a man eating fish? A seafood restaurant

[Explanation: The man can be eating a fish, or the fish can be a man-eating one.]

9) How do you make a hot dog stand? Steal its chair

[Explanation: Can refer to a stand where they sell hot dogs, or a giant hot dog standing on two legs.]

10) What did the clerk say when the woman asked to try on the dress in the window? Don't you think it would be better to use the dressing room?

[Explanation: Is the woman going to the dressing room to try on the dress from the window or standing in the window to try on a dress?]

11) Will you join me in a bowl of soup? Do you think there's room for both of us?

[Explanation: "Join me in" can mean eat with me or physically get into with me.]

12) Why did the bear tiptoe through the campground? He didn't want to wake up the sleeping bags.

[Explanation: "Sleeping bags" can be either bags that are sleeping or bags that are FOR sleeping.]

13) What kind of stamp do you have to stick on yourself? None. You stick them on envelopes.

[Explanation: Is the stamp being stuck by you, or are you sticking it onto a part of your body?]

14) What did the doctor say to the patient who thought he was getting smaller? You'll just have to be a little patient.

[Explanation: "Patient" can refer to the person seeing the doctor, or to a way of acting.]

Okay! Great job with those. Now let's try writing some of our own. First we choose a topic. Do any of you play sports? [Suggest baseball.]

First, we need to make a list of words that have to do with _____. Can anybody suggest some terms you use when you play ____ ?

[Baseball terms: bat, ball, pitcher, catcher, plate, base, homerun, fly ball, shortstop, umpire, run, slide, fielder, etc.]

Good! Now let's look at your list and see if any of these words are homonyms. Do you remember what a homonym is? If any of the words have more than one meaning, maybe we can make it into a riddle...

[Baseball riddles:

BALL:

Why did Cinderella get thrown off the baseball team? Because she ran away from the ball.

PITCH:

Why do you need a baseball player with you when you go camping? To pitch the tent.

STOLEN BASE:

Why did the cop run across the baseball field? Someone stole second base.

BATTER:

Why is a cake like a baseball team? They both need a good batter.

FLIES (FLY BALLS):

Why are spiders good baseball players? Because they are good at catching flies.

SHORTSTOP:

Does it take longer to run from 1st base to 2nd base or 2nd base to 3rd base? From 2nd to 3rd, because there is a shortstop in the middle.

DIAMOND:

Where is the largest diamond in New York City? Yankee Stadium (or Shea Stadium).

PLATE:

Where does a catcher sit for dinner? Behind the plate.

Why do umpires make the best dinner guests? Because they are always cleaning the plate.

After completing list of model riddles, assign pairs to work together to create riddles on a new topic. Give them a notebook in which to write and illustrate the riddles.]

See if you can think of any more riddles to bring me next time we meet!

Session Four

Today we're going to read Amelia Bedelia books. Do you know her? The first book is called *Amelia Bedelia and the Surprise Shower* by Peggy Parish.

Here's what I want you to do. I want you to read the book out loud. Read as you normally would, but make sure you think about the meaning while you read. Stop every time you come to a sentence you think might have more than one meaning. We're going to write down all of the sentences with more than one meaning on this chart.

Do you understand what I want you to do?

[Child begins reading. Whenever a child stops on an ambiguous sentences, ask him/her to explain how Amelia understands the sentence, as well as how the sentence is meant to be understood. Keep a chart of the ambiguities and the child's interpretations of the ambiguous utterances.

Ambiguities:

p. 6 "shower"

pp.10-12 "ice" the cupcakes and the fish

pp.14-15 "prune the hedge"

p.18 "scale" the fish

p.22 "run over the tablecloth with an iron"

p.30 "cut" the flowers]

Amelia Bedelia's Family

Great! The next book is called *Amelia Bedelia's Family Album* by Peggy Parish.

[Show child edited book, containing 7 pages from *Family Album*, which chronicles what all of the members of Amelia's family do. The ambiguous pages include:

p.11 “big-game hunter”
 p.13 “bank teller”
 p. 15 “boxer”
 p. 21 “takes pictures”
 p. 31 “works with Clay”
 p.33 “garbage collector”
 p. 35 “belongs to a fan club”
 p. 37 “stuffs olives”

Let’s read it the same way we did the last book. [Have children read the first few pages: boxer, bank teller, and big-game hunter. Stop them for explanations along the way.]

Great job! But look, the book is incomplete! We’re going to finish the book, the way you think Amelia might have.

[The resolution side of the page is now covered. Have child read the front of the page aloud, then prompt him/her for what Amelia might have said. Let child illustrate the page.

For example, the available side of the page might say:

“Uncle Dan takes pictures,” said Amelia Bedelia.

“What kind of pictures does he take?” asked Mr. Rogers.

Participants now turn the page and write/draw the answer Amelia will provide—that he is a thief.

This portion contains: Cook, takes pictures, garbage collector, and fan club]

Way to go. Now we’re going to write in a few relatives of our own. What if Amelia had a cousin who was a ... ? [Suggest ambiguously-titled profession, have children draw from AB’s point of view. Professions to include: waiter, train conductor, mail carrier, taxi driver.]

Let’s talk about what you’ve learned working with me. [Let children talk, but steer conversation toward need to reprocess some language and strategies for comprehension monitoring like re-reading confusing passages, looking for alternate definitions for words, and checking to see if what you understand fits with the rest of the text.]

Appendix J

CONTROL GROUP SCRIPT

Mouse Soup by Arnold Lobel

My name is Marcy. Remember the last time we met I told you that I'm trying to figure out what is the best way to teach children how to read? Well, I'm going to be seeing some of the children in your class, but not all, for the next few weeks, to see if I can help them understand what they read a little bit better.

Here is what is going to happen: we are going to meet once a week and we're going to read this book [show MOUSE SOUP]. Then we will talk about it. I expect you to try your best and think really hard about everything we do. Can you all do that? ... Good.

Before we read the book, let's talk about the different parts of the book. What is the first thing you read when you start a new book? [TITLE] Now we want to look to see who wrote the book—do you know what we call the person who wrote the book? [AUTHOR, Arnold Lobel] Good. One of the reasons you might want to look at the author's name is because you may have read other things that author wrote. Sometimes if you liked one thing an author wrote, you find you like other books that author writes. Who is your favorite author?

Okay, let's open the book. This first page, with the title of the book on it, is called the title page—makes sense right? But we already know the title, so the next page we need to read is this one. Do you know what this page is called? [TABLE OF CONTENTS] Good work! Can you explain how a table of contents page works? [Agree that the new stories start on the pages listed, and that if we only wanted to read a specific story, we could flip directly to that page.]

Can you tell me, just from looking at the table of contents, how long the introduction is? [Explain that if the first story begins on p.12, the intro must be 11 pages... but oops—look it starts on p.6! $11-6=5$] Okay, so now that we know the introduction is five pages long, let's figure out how long each story is. [Subtract page numbers.] Good job! Now we're going to read the book out loud. Why don't you start by reading the introduction?

Great job reading that book! Next time I see you, we will talk about what we read in Mouse Soup.

Next Session

Last time we met, we read the book *Mouse Soup* by Arnold Lobel. Do you remember what it was about? [Let kid talk for a few minutes to refresh their memory of the book.] We're going to make a chart today to help us work out everything about the book. These are the big things I want to talk about plot, point of view, setting, and imagery. [Write these on white board.]

First, let's define what all of these things mean. Do you know what ____ means?

PLOT: actions or events of a story

POV: a way of looking at things

SETTING: where the action took place

IMAGERY: a picture in the mind of something not real, or not present at the time

Good! Now let's talk about the plot of *Mouse Soup*. What were the main actions or events of *Mouse Soup*? [Distinguish between the plot of the book and the plots of the mouse's stories. Write big thoughts down, or have children write down. Encourage the children to look back through the book for answers throughout.]

Whose POV did we see the actions in *Mouse Soup* from? [The main plot was through the mouse's eyes, but the stories are varied...]

Where is the book set? [Weasel's house, mouse's imagination, various locations...]

What kind of images does the mouse create with his stories? How do the illustrations help? Did you imagine anything different from what you see in the illustrations? [If there's time, encourage the children to illustrate other parts of the book.]

Continue reading from book and discussing from chart at the beginning of each session.

References

- Abrahamsen, E.P., & Shelton, K.C. (1989). Reading comprehension in adolescents with learning disabilities: semantic and syntactic effects. *Journal of Learning Disabilities, 22*, 570-572.
- August, D.L., Flavell, J.H., & Clift, R. (1984). Comparison of comprehension monitoring of skilled and less skilled readers. *Reading Research Quarterly, 20*, 39-53.
- Baker, L. (1979). Comprehension monitoring: Identifying and coping with text confusions. *Journal of Reading Behavior, 11*, 365-374.
- Baker, L. (1984). Spontaneous versus instructed use of multiple standards for evaluating comprehension: Effects of age, reading proficiency, and type of standard. *Journal of Experimental Child Psychology, 38*, 289-311.
- 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. 155-200). London: Academic Press.
- Baker, L. (1989). Metacognition, comprehension monitoring and the adult reader. *Educational Psychology Review, 1*, 3-38.
- Baker, L. & Anderson, R.I. (1982). Effects of inconsistent information on text processing: Evidence for comprehension monitoring. *Reading Research Quarterly, 17*, 281-294.
- Baker, L. & Zimlin, L. (1989). Instructional effects on children's use of two levels of standards for evaluating their comprehension. *Journal of Educational Psychology, 81*, 340-346.
- Bernstein, J. (1979). *Fiddle with a riddle*. New York: Dutton.
- Bohannon, J.N. (1975). The relationship between syntax discrimination and sentence imitation in children. *Child Development, 46*(2), 444-451.
- Bowey, J.A. (1986). Syntactic awareness in relation to reading skill and ongoing reading comprehension monitoring. *Journal of Experimental Child Psychology, 41*, 282-299.
- Bryant, P.E., & Bradley, L. (1985). *Children's reading problems: Psychology and Education*. Oxford, England: Basil Blackwell.
- Cairns, H. C. (1999). *Psycholinguistics: An introduction*. Austin, TX: Pro-Ed.

Cairns, H.S, Waltzman, D., & Schlisselberg, G. (2004). Detecting the ambiguity of sentences: Relationship to early reading skill. *Communication Disorders Quarterly*, 25(2), 68-78.

Carlisle, J. (1995). Morphological awareness and early reading achievement. In L. Feldman (Ed.), *Morphological aspects of language processing* (pp. 189-209). Hillsdale, NJ: Lawrence Erlbaum Associates.

Catts, H.W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research*, 36, 948-958.

Cornoldi, C. (1991). Metacognitive control processes and memory deficits in poor comprehenders. *Learning Disabilities Quarterly*, 13, 245-256.

Cornoldi, C., & Oakhill, J.V. (1996). *Reading comprehension difficulties: Processes and interventions*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cosgrove, J.M., & Patterson, C.J. (1978). Generalization of training for children's listener skills. *Child Development*, 49(2), 513-516.

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(3), 277-299.

Dewey, J. (1910). *How we think*. Boston: Heath.

Downing, J. (1978, Nov 30-Dec 2). The psycholinguistic basis of cognitive clarity. Paper presented at the National Reading Conference, St. Petersburg Beach, FL.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary, Third Edition*. Circle Pines, MN: AGS Publishing.

Durkin, D. (1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 14(4), 481-533.

Ehri, L.E. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. Metsala and L. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3-40). Mahwah, NJ: Erlbaum.

Ericsson, K.A. & Simon, H.A. (1980). Verbal reports as data. *Psychological review*, 87(3), 215-251.

Fowler, A.E. (1988). Grammaticality judgments and reading skill in grade 2. *Annals of Dyslexia*, 38, 73-94.

- Fowles, B., & Glanz, M.E. (1977). Competence and talent in verbal riddle comprehension. *Journal of Child Language*, 4, 433-452.
- Gagne, R.M. (1977). *The conditions of learning (3rd ed.)*. New York: CBS College Publishing.
- Gaux, C., & Gombert, J.E. (1999). Implicit and explicit syntactic knowledge and reading in pre-adolescents. *British Journal of Developmental Psychology*, 17, 169-188.
- Glass, A.L., & Perna, J. (2001). The role of syntax in reading disability. *Journal of Learning Disabilities*, 19(6), 345-359.
- Gombert, J.E. (1992). *Metalinguistic development*. Chicago: University of Chicago Press.
- 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*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gough, P.B., & Tunmer, W.E. (1986). Decoding, reading and reading disability. *Remedial and Special Education*, 7(1), 6-10.
- Hacker, D.J. (1997). Comprehension monitoring of written discourse across early-to-middle adolescence. *Reading and Writing: An Interdisciplinary Journal*, 9(3), 207-240.
- Harris, A.J., & Sipay, E.R. (1990). *How to increase reading ability: A guide to developmental and remedial methods*. New York: Longman.
- Harris, P.L., Kruithof, A., Terwogt, M., & Visser, T. (1981). Children's detection and awareness of textual anomaly. *Journal of Experimental Child Psychology*, 31, 212-230.
- Hirsh-Pasek, K., Gleitman, L.R., & Gleitman, H. (1978). What did the brain say to the mind? A study of the detection and report of ambiguity by young children. In A. Sinclair, R.J. Jarvella, & W.J.M. Levelt, (Eds.), *The child's conception of language* (pp. 97-132). New York: Springer-Verlag.
- Ironsmith, M., & Whitehurst, G.J. (1978). The development of listener abilities in communication: How children deal with ambiguous information. *Child Development*, 49, 348-352.
- Isakson, R.L., & Miller, J.W. (1976). Sensitivity to syntactic and semantic cues in good and poor comprehenders. *Journal of Educational Psychology*, 68, 787-792.

Karlsen, B., Madden, R., & Gardner, E.F. (1976). *Stanford Diagnostic Reading Test Manual*. New York: Harcourt Brace Jovanovich, Inc.

Kazemek, F. (1999). Why was the elephant late in getting on the ark? Elephant riddles and other jokes in the classroom. *The Reading Teacher*, 52(8), 896-898.

Kennedy, D., & Weener, P. (1973) Visual and auditory training with the cloze procedure to improve reading and listening comprehension. *Reading Research Quarterly*, 8, 524-541.

Kinnunen, R., & Vaurus, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction*, 5, 143-165.

Kintsch, W., & van Dijk, T.A. (1978) Toward a model of discourse comprehension and production. *Psychological Review*, 83, 363-394.

Kuhn, M., & Stahl, S. (1998). Teaching children to learn word meanings from context: A synthesis and some questions. *Journal of Literacy Research*, 30, 119-138.

Layton, A., Robinson, J., & Lawson, M. (1998). The relationship between syntactic awareness and reading performance. *Journal of Research in Reading*, 21(1), 5-23.

Leong, C.K. (1984). Cognitive processing, language awareness, and reading in grade 2 and grade 4 children. *Contemporary Educational Psychology*, 9, 369-383.

Lobel, A. (1979). *Frog and toad are friends*. New York: HarperCollins Children's Books.

Lobel, A. (1977). *Mouse soup*. New York: HarperCollins Children's Books.

Lundberg, I., Frost, K., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly*, XXIII(3), 263-284.

MacGinitie, W. H., MacGinitie, R.K., Maria, K. & Dreyer, L. (2000). *Gates-MacGinitie reading tests (4th edition)*. Chicago: Riverside Publishing Company.

Mahony, D.L., & Mann, V.A. (1992). Using children's humor to clarify the relationship between linguistic awareness and early reading ability. *Cognition*, 45, 163-186.

Mahony, D.L., & Mann, V.A. (1998). Children's linguistic humor and reading research: A match made in heaven. *Cahiers de Psychologie Cognitive*, 17(2), 287-311.

Markman, E.M. (1977). Realizing you don't understand: A preliminary investigation. *Child Development, 46*, 986-992.

Markman, E.M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development, 50*, 643-655.

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

McKeown, M. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly, 20*, 482-496.

Nagy, W., & Scott, J. (2000). Vocabulary processes. In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.) *Handbook of reading research: Volume III* (pp. 269-284). Mahwah, NJ: Erlbaum.

National Institute of Child Health and Human Development (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: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Neale, M.D. (1966). *The Neale Analysis of Reading Ability* (2nd ed.). London: Macmillan Education.

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

Oakhill, J.V. (1994). Individual differences in children's test comprehension. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*, pp. 821-849. New York: Academic Press.

O'Donnell, M.P. & Wood, M. (1989). *Informal Assessment of Elementary Reading*. Reading Associates.

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

Parish, P. (1979). *Amelia Bedelia and the surprise shower*. New York: HarperCollins.

Parish, P. (1985). *Amelia Bedelia goes camping*. New York: Avon Books.

Parish, P. (1988). *Amelia Bedelia's family album*. New York: Greenwillow Books.

- Patterson, C.J., O'Brien, C., Kister, M.C., Carter, D.B., & Kotsonis, M.E. (1981). Development of comprehension monitoring as a function of context. *Developmental Psychology*, *17*, 379-389.
- Pea, R.D. (1980). The development of negation in early child language. In D. R. Oldon (Ed.), *The social foundations of language and thought: Essays in honour of Jerome S. Bruner* (pp. 156-186). New York: Norton.
- Perfetti, C.A., & McCutchen, D. (1986). Schooled language competence: Linguistic abilities in reading and writing. In S. Rosenberg (Ed.), *Advances in Applied Psycholinguistics*. Cambridge University Press: Cambridge.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT/Bradford Books.
- Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.), *Handbook of reading research: Volume III* (pp. 545-561). Mahwah, NJ: Erlbaum.
- Rinehart, S.D., & Platt, J.M. (1984). Metacognitive awareness and monitoring in adult and college readers. *Forum for Reading*, *56*, 654-663.
- Revelle, G.L., Wellman, H.M., & Karabenick, J.D. (1985). Comprehension monitoring in preschool children. *Child Development*, *56*, 654-663.
- Rosenshine, B. & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, *64*(4), 479-530.
- Roth, F.P., Speece, D.L., Cooper, D.H., & De La Paz, S. (1996). Unresolved mysteries: How do metalinguistic and narrative skills connect with early reading? *The Journal of Special Education*, *30*(3), 257-77.
- Ruffman, T. (1993, September). *Internal consistency: When do children come to understand that claims or beliefs should be logically consistent?* Paper presented at the British Psychological Society Annual Meeting, Birmingham.
- Ruffman, T. (1995). *Constructive processing and comprehension monitoring*. Unpublished manuscript, University of Sussex, Brighton.
- Ruffman, T. (1996). Reassessing children's comprehension-monitoring skills. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention*. (pp. 33-67). Mahwah, NJ: Erlbaum Associates.
- Ruffman, T., Billins, Z., & White, D. (1995). *Another look at children's comprehension-monitoring skills*. Unpublished manuscript, University of Sussex, Brighton.

- Rumelhart, D. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance: Vol. 6* (pp. 573-603). New York: Academic Press.
- Ryan, E., & Ledger, G. (1979). Grammaticality judgments, sentence repetitions, and sentence corrections of children learning to read. *International Journal of Psycholinguistics*, 6, 23-40.
- Ryan, E., & Ledger, G. (1984). Learning to attend to sentence structure: Links between metalinguistic development and reading. In J. D. R. Valtin (Ed.), *Language awareness and learning to read* (pp. 149-171). Berlin: Springer-Verlag.
- Scholl, D.M. & Ryan, E.B. (1980). The development of metalinguistic performance during the early school years. *Language and Speech*, 23, 199-211.
- Schooler, J.W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology, General*, 122(2), 166-183.
- Shankweiler, D., Crain, S., Brady, S., & Macaruso, P. (1992). Identifying the causes of reading disability. In D.B. Gough, L.C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 275-305). Hillsdale, NJ: Erlbaum.
- Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K., Fletcher, J., Brady, S., Fowler, A., Dreyer, L., Marchione, K., Shaywitz, S., & Shaywitz, B. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. *Scientific Studies of Reading*, 3, 69-94.
- Shultz, T.R. (1974). Development of the appreciation of riddles. *Child Development*, 45, 100-105.
- Shultz, T.R. & Pilon, R. (1973). Development of the appreciation of verbal jokes. *Developmental Psychology*, 10, 13-20.
- Slingerland, B. (1977). *Revised pre-reading screening procedures*. Cambridge, MA: Educators Publishing Service.
- Snow, C.E., Burns, S.M., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stanovich, K.E. (1992). Speculation on the causes and consequences of individual differences in reading acquisition. In P. Gough et al. (Eds.), *Reading Acquisition* (pp.307-342). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-659.

Swinney, D.A., Prather, P., & Love, T. (2000). The time-course of lexical access and the role of context: Converging evidence from normal and aphasic processing. In Grodzinsky, Yosef, Shapiro, Lewis P., & Swinney, David (Eds.), *Language and the brain: Representation and processing* (pp 273-292). San Diego, CA: Academic Press.

Thorndike, E.L. (1917). Reading as reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology*, 8, 323-332.

Tunmer, W. E., & Bowey, J.A. (1984). Metalinguistic awareness and reading acquisition. In C. Pratt, W.E. Tunmer, & M.L. Herriman (Ed.), *Metalinguistic awareness in children*. New York: Springer.

Tunmer, W. E., Herriman, M.L., & Nesdale, A.R. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly*, 23(2), 134-158.

Tunmer, W.E., Nesdale, A.R., & Wright, A.D. (1987). Syntactic awareness and reading acquisition. *British Journal of Developmental Psychology*, 5, 25-34.

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

Wagoner, S.A. (1983). Comprehension monitoring: What it is and what we know about it. *Reading Research Quarterly*, 18, 328-346.

Weaver, P. (1979). Improving reading instruction: effects of sentence organization instruction. *Reading Research Quarterly*, 15, 127-146.

Weinstein, C.E. (1978). Teaching cognitive elaboration learning strategies. In H.F. O'Neal (Ed.). *Learning strategies*. New York: Academic Press.

Werner, H., & Kaplan, E. (1952). The acquisition of word meanings: A developmental study. *Monographs of the Society of Research in Child Development*, 15(Serial No.1).

Woodcock, R. W. (1973). *Woodcock reading mastery test revised (WRMT-R)*. Circle Pines, MN: American Guidance Service.

Woodcock, R. W. (1987). *Woodcock reading mastery test revised (WRMT-R)*. Circle Pines, MN: American Guidance Service.

Woodcock, R. W. (1998). *Woodcock reading mastery tests revised normative update (WRMT-R/NU)*. Circle Pines, MN: American Guidance Service.

Yalisove, D. (1978). The effect of riddle structure on children's comprehension of riddles. *Developmental Psychology*, *14*, 173-180.

Yuill, N. (1996). A funny thing happened on the way to the classroom: Jokes, riddles, and metalinguistic awareness in understanding and improving poor comprehension in children. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention*. Mahwah, NJ: Lawrence Erlbaum.

Yuill, N. (1998). Reading and riddling: The role of riddle appreciation in understanding and improving poor text comprehension in children. *Cahiers de Psychologie Cognitive*, *17*(2), 313-342.

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

Zabucky, K., & Ratner, H. H. (1986). Children's comprehension monitoring and recall of inconsistent stories. *Child Development*, *57*, 1401-1418.

Zabucky, K., & Ratner, H.H. (1992). Children's comprehension monitoring and recall of inconsistent stories. *Child Development*, *57*, 1401-1418.

Zipke, M. (2005, in press). Metalinguistic awareness and reading comprehension. *Reading psychology*.