

ABSTRACT

Title of dissertation: READING COMPREHENSION COMPONENT
 PROCESSES IN EARLY ADOLESCENCE

 Jennifer Grace Cromley, Doctor of Philosophy, 2005

Dissertation directed by: Assistant Professor Roger Azevedo
 Department of Human Development

A significant proportion of American high school students struggle with reading comprehension. Several different models might help identify the components that have the largest effect on comprehension. The current dissertation study replicates a comparison of the Construction-Integration (CI), Verbal Efficiency (VE), and Inferential Mediation (IM) models of reading comprehension, the latter model based on an extensive literature review. It then tests the fit of four variations on the IM model.

Ninth-grade students ranging from 1st to 99th percentile on comprehension completed measures of background knowledge, inferencing, strategies, vocabulary, word reading and comprehension. Researcher-developed measures of background knowledge, inferencing and strategies (based on Cromley & Azevedo, 2004a) showed good reliability with this sample.

A subset of the students also completed a think-aloud protocol while reading a passage from an American history textbook. These protocols were transcribed and coded

using a coding scheme adapted from Azevedo, Guthrie, and Seibert (2004).

As in a preliminary study, the IM model had a much better fit to the data than did the CI or VE models. The original IM Model had the best fit, explaining 66% of the variance in comprehension. All predictors made a significant contribution to comprehension, with vocabulary, background knowledge, and strategies having significant indirect effects. Vocabulary and background knowledge made the greatest total contribution to comprehension. There were large, significant differences between low- and high-comprehending participants on all of the predictor variables, except for word reading accuracy, where there were small but significant differences.

The coded think-aloud protocols were largely consistent with the correlations underlying the model. Spearman rank correlations among the codes provide convergent evidence for eleven of the correlations underlying the model. The think-aloud protocols also provided convergent evidence for the validity of the paper-and-pencil measures.

The current study validates and refines a new model of reading comprehension. Results suggest that both the direct and indirect effects of the components are important for comprehension. Results also suggest that vocabulary and background knowledge might first be targeted for interventions with 9th grade students who struggle with reading comprehension. Implications for future research are also discussed.

READING COMPREHENSION COMPONENT PROCESSES
IN EARLY ADOLESCENCE

by

Jennifer Grace Cromley

Dissertation submitted to the Faculty of the Graduate School
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Committee:

Assistant Professor Roger Azevedo, Chair
Professor Mariam Jean Dreher
Professor John T. Guthrie
Professor Gregory Hancock
Professor Allan Wigfield

© Copyright by
Jennifer Grace Cromley
2005

ACKNOWLEDGMENTS

Thanks to the following people and organizations, listed in chronological order of their appearance in my life: Dr. Elizabeth C. Cromley, my mother, for inspiring, questioning, guiding, and editing; St. Ann's School, Brooklyn, NY; the late Joseph D. Ernst, who always believed in me; Catherine White, my partner in love and life; Dr. Roger Azevedo, for mentoring, teaching, editing, encouraging, and for help with pilot data collection and coding, and dissertation research materials, equipment, supplies, and data analysis; a University of Maryland College Park Graduate School fellowship; Dr. Robert J. Mislevy for encouraging me to look at psychological models of reading; the AERA/Spencer Pre-Dissertation Fellowship program; my dissertation committee—Dr. Roger Azevedo, Dr. M. Jean Dreher, Dr. John T. Guthrie, Dr. Gregory Hancock, and Dr. Allan Wigfield—for asking the tough questions; Dr. Melanie Kuhn for advice on measuring fluency; Cara Sanderson for assistance with data entry; Dr. Ana Taboada for assistance with defining self-questioning; the newest member of the family, Bayard Cromley Lipscomb White; a Spencer Foundation Dissertation Fellowship; Dr. Hunter Hogewood for re-coding and establishing interrater reliability on the free recall protocols; Dan Moos for re-coding and establishing interrater reliability on the think-aloud protocols; and Kelly Ryan for historical fact-checking.

TABLE OF CONTENTS

List of Tables	viii
List of Figures	x
CHAPTER I: RATIONALE	1
Reading Proficiency of and Reading Demands on Adolescents	2
Need for Basic Research in Adolescent Literacy	4
Prior Research on Components of Reading Comprehension in High School	4
Single-variable studies	5
Multi-component studies	6
Think-aloud studies	6
Coordinating product and process data	7
The Inferential Mediation Model	8
The Construction-Integration Model	9
Evidence for the model	13
Verbal Efficiency Theory	14
Evidence for the theory	17
Differences Between the IM, CI, and VE Models	18
Building on the Preliminary Study	21
The Present Study	22
Research Questions	22
CHAPTER II: REVIEW OF THE LITERATURE	23
Path Analysis	23
Building the Inferential Mediation (IM) Model	26
Stage 1: Identifying Predictor Variables for the IM Model	27
Definitions of the variables	27
Stage 2: Evidence for the IM Model	32
Criteria for selecting studies	32
The IM Model	33
Path 1: Effect of background knowledge on comprehension	36
Path 2: Effect of background knowledge on strategies	47
Path 3: Effect of background knowledge on inference	52
Path 4: Effect of strategies on comprehension	56

Chapter II (Cont'd.)	
Path 5: Effect of strategies on inference	62
Path 6: Effect of word reading on comprehension.....	64
Path 7: Effect of vocabulary on inference.....	73
Path 8: Effect of vocabulary on comprehension.....	75
Path 9: Effect of inference on comprehension.....	83
Path 10: Effect of word reading on vocabulary	91
Path 11: Correlation between background knowledge and vocabulary.....	94
Path 12: Correlation between background knowledge and word reading	94
Summary of the IM model.....	95
Variations on the IM Model.....	95
Coordinating Product and Process Data	97
Science and mathematics	98
Reading	103
Statement of the Research Problem	111
Research Questions.....	113
CHAPTER III: PRELIMINARY STUDY.....	115
Rationale	115
Method	115
Participants.....	115
Research Design.....	116
Materials and Measures	116
Word reading	116
Background knowledge	117
Vocabulary.....	117
Reading comprehension.....	118
Inferences.....	118
Reading strategies	119
Think-aloud materials.....	119
Equipment.....	120
Procedure	120
Component measures.....	121
Think-aloud.....	123
Data Analysis and Scoring.....	124
Scoring procedures for the component measures	124
Think-aloud data.....	125
Coding scheme for the think-aloud protocols.....	125
Interrater Reliability.....	128
Results	129
Reading Components—Descriptive Analyses.....	129
Path Analysis	131
Think-Aloud Data.....	137
Results of the Spearman rank correlation analyses on the think-aloud data.....	139
Discussion.....	141
Implications for the Dissertation Study	142

CHAPTER IV: METHOD.....	144
Participants.....	144
Research Design.....	146
Materials and Measures	146
Word reading	146
Word reading accuracy	146
Word reading fluency	147
Background knowledge	147
Vocabulary.....	148
Reading comprehension.....	149
Inferences.....	149
Reading strategies	150
Think-aloud materials	150
Equipment.....	151
Procedure	151
Component measures	152
Think-aloud.....	153
Data Analysis.....	155
Scoring procedures for the component measures	155
Principal component scores for word reading	156
Think-aloud and verbal recall data	157
Coding scheme for the think-aloud protocols.....	157
Verbal recall protocols.....	161
Path analyses.....	161
Inter-rater agreement.....	162
CHAPTER V: RESULTS.....	164
Descriptive Analyses	164
Reading components measures.....	164
Evidence for validity of the measures.....	173
Research Question 1	173
Research Question 2	174
Research Question 3	177
Research Question 4	178
Research Question 5	181
Descriptive results of the think-aloud data	181
Descriptive results of the verbal recall data.....	188
Results of the Spearman rank correlation analyses on the think-aloud data.....	192
Examples from the think-alouds corresponding to paths in the model.....	197

CHAPTER VI: DISCUSSION	200
Sample and Think-Aloud Sub-Sample	200
Data screening.....	201
Model Comparisons	201
Replicating the comparison of the CI, VE, and IM Models	201
Refining the IM model by testing four variations.....	202
Relative Contribution of the Components to Reading Comprehension.....	206
Differences Between High- and Low-comprehending Readers	217
Limitations	222
Implications.....	224
Theoretical contributions	224
For future research	225
Methodological contributions	225
Practical contributions	226

Appendices

Appendix A: Comprehension Strategy Instruction Studies Showing Effects on Reading Comprehension With Middle- and High-School Students, Listed in Order by Strategy	227
Appendix B: Strategies Enacted by Middle and High School Readers in Think-Aloud Studies, Listed Alphabetically by Author	240
Appendix C: Preliminary Study Parent Cover Letter, Parental Consent Form, Student Assent, and Student Background Information Sheet	245
Appendix D: Preliminary Study Background Knowledge Test (from Cromley & Azevedo, 2004a)	250
Appendix E: Preliminary Study Inference Test (from Cromley & Azevedo, 2004a)	254
Appendix F: Preliminary Study Think-Aloud Practice Text (adapted from Roller, 1986)	259
Appendix G: Think-Aloud Protocol Text (from Viola, Wheatley, & Hart, 1998)	262
Appendix H: Preliminary Study Think-Aloud Codes, Definitions, and Examples (adapted from Azevedo, Guthrie, & Seibert, 2004).....	266
Appendix I: Dissertation Study Parent Cover Letter, Parental Consent Form, Student Assent, and Student Background Information Sheet	271
Appendix J: Dissertation Study Passage used for one-minute fluency measure (Leslie & Caldwell, 2000)	276

Appendix K: Dissertation Study Background Knowledge Test (adapted from Cromley & Azevedo, 2004a)	278
Appendix L: Dissertation Study Selected Items from the Vocabulary Measure	281
Appendix M: Dissertation Study Inference and Strategy Use Test (adapted from Cromley & Azevedo, 2004a)	284
Appendix N: Dissertation Study Think-Aloud Practice Text (adapted from Roller, 1986)	293
Appendix O: Dissertation Study Think-Aloud Codes, Definitions, and Examples (adapted from Azevedo, Guthrie, & Seibert, 2004)	295
Appendix P: Dissertation Study Verbal Recall <i>a priori</i> Coding Rubric	300
References Cited	303

LIST OF TABLES

1.	Number of Experimental Studies Establishing Each Path and Corroborating Studies.....	35
2.	Prior Knowledge Strategy Instruction Interventions Showing Effects on Reading Comprehension, Ranked by Age	39
3.	Reported Correlations between Background Knowledge and Reading Comprehension, Ranked by Age	46
4.	Reported Correlations between Background Knowledge and Strategy Use, By Author	51
5.	Reported Correlations between Strategies and Comprehension, By Author	61
6.	Reported Correlations between Word Reading and Comprehension, By Age (Studies not Included in the Gough et al. [1996] Meta-Analysis).....	71
7.	Reported Correlations between Vocabulary and Reading Comprehension, by Author	82
8.	Reported Correlations between Inference and Reading Comprehension, by Age	89
9.	Reported Correlations between Word Reading and Vocabulary, By Author	92
10.	Descriptive Statistics for Reading Comprehension and Component Measures...130	
11.	Correlations and Reliabilities for Reading Measures	132
12.	Variance/Covariance Matrix for Reading Comprehension and Predictor Component Measures	133
13.	Indicators of Fit for the Three Reliability-Adjusted Models	134
14.	Direct and Indirect Paths from Predictors to Reading Comprehension, for the CI, VE, and IM Models.....	136
15.	Proportion of Use for Each Code in the Think-Aloud Coding Scheme, Across Groups.....	137

16.	Spearman Rank Correlations Among Proportions of Verbalization for the Five Predictor Variables	140
17.	Demographics for the Whole Sample and the Think-Aloud Sub-Sample	144
18.	Descriptive Statistics on the Measures for the Entire Sample and the Think-Aloud Sub-Sample	165
19.	Sources of Variance and Intraclass Correlations for Class and Teacher Effects	167
20.	Descriptive Statistics and Results of T-tests Comparing the Think-Aloud Sample to Participants who Were not Selected for Think-Alouds	168
21.	Correlations and Reliabilities for the Measures.....	170
22.	Fit Indices for the CI, VE, and IM Models	174
23.	Fit Indices for the Four Variations on the IM Model.....	175
24.	Standardized Direct and Indirect Paths from Predictors to Reading Comprehension in Model 2.....	177
25.	Group Means, Standard Deviations, and Results of Independent Sample T-tests for Predictors and for Comprehension Across Low- and High-Comprehending 9 th Grade Students	179
26.	Raw Frequency and Proportion of Use for Each Code in the Think-Aloud Coding Scheme, Across Groups	183
27.	Spearman Rank Correlations Between Scores on Each Paper-and-Pencil Revolutionary War Measure and the Proportion of Verbalizations in the Corresponding Think-Aloud Variable	191
28.	Spearman Rank Correlations Among Proportions of Accurate Verbalization of the Predictor Variables and Verbal Recall.....	192
29.	Spearman Rank Correlations Among Proportions of Inaccurate Verbalization of the Predictor Variables and Verbal Recall.....	193
30.	Spearman Rank Correlations Between Proportions of Accurate and Inaccurate Verbalization of the Five Predictor Variables	194
31.	Summary of Significant Correlations among the Predictor Variables from the Paper-and-Pencil and Think-Aloud Data.....	196

LIST OF FIGURES

1. Path Diagram for the Construction-Integration Model.....12

2. Path Diagram for the Verbal Efficiency-Based Model.....17

3. The Inferential Mediation Model, Construction-Integration Model, and Verbal Efficiency Models.....20

4. Illustrative Path Diagram26

5. Path Diagram for the Inferential Mediation Model 34

6. The Four Variations of the Inferential Mediation Model96

7. Final Solution for the Inferential Mediation Model.....135

8. Final Solution for Model 2.....176

CHAPTER I: RATIONALE

Many young adolescents struggle with reading comprehension of academic texts—25% of American 8th grade students performed at the “Below Basic” level on the National Assessment of Educational Progress reading tests in 2002 (Grigg, Daane, Jin, & Campbell, 2003). Reading is a critical academic skill, one which is necessary for success in all academic domains. However, the empirical research base on adolescents who struggle with reading comprehension is very small (see Curtis, 2002; Underwood & Pearson, 2004), and the quantitative portion of that research base is even smaller.¹ We need to understand what young adolescents who struggle with reading comprehension struggle with in order to in design future effective educational interventions.

I have therefore conducted a study measuring various components of academic reading comprehension for students in 9th grade across a range of reading abilities that converges two sources of data: closed-ended measures and think-aloud protocols. Four variations of a new model of reading comprehension for academic texts were fit to the data, low- and high-comprehending students were compared on the comprehension components, and significant direct and indirect paths from the model were sought in the think-aloud protocols. Implications for theory, future research in basic processes and interventions, and teaching are then discussed.

¹ Only quantitative studies were reviewed because they are recommended for building generalizable interventions.

Reading Proficiency of and Reading Demands on Adolescents

A large, but stable, proportion of young adolescent students struggle with reading comprehension. On the 2002 NAEP reading tests, 25% of eighth-grade students performed at the Below Basic level; 43% performed at the “Basic” level; only 30% performed at the “Proficient” level; and 3% at the “Advanced” level (Grigg et al., 2003). The Proficient level is the goal for all students as defined by the NAEP governing body (National Assessment Governing Board [NAGB], 1999), but was met or exceeded by only 33% of eighth-grade students in 2002. In eighth grade, students at the Proficient level are “able to show an overall understanding of the text, including inferential as well as literal information to extend the ideas in the text by making clear inferences from it, by drawing conclusions, and by making connections to their own experiences” (NAGB, p. 35). At the Basic level, students “should demonstrate a literal understanding of what they read and be able to make some interpretations” (NAGB, p. 34); at the Below Basic level, students do not even demonstrate this level of performance. At the Advanced level, students “describe the more abstract themes and ideas of the overall text” (NAGB, p. 35). NAEP is the largest single study of adolescent reading in the United States, involving approximately 115,000 eighth-grade students from 45 states and 5 jurisdictions in the 2002 administration (Grigg et al., 2003). Other large-scale studies show similar proportions of young American adolescents who struggle with reading comprehension (Brown & Fetters, 1984; Owings, 1995).

Young adolescents’ performance on NAEP reading tests have remained quite stable over the last 30 years: average scores for the middle two quartiles for 8th grade students were 258 in 1971 and 1984, and 261 in 1999; average scores for the lowest quartile were 212 in 1971, 215 in 1984, and 214 in 1999; and average scores in the

highest quartile were 293 in 1971, 296 in 1984, and 302 in 1999 (National Center for Education Statistics, 2000)². Over this time period, the proportion of school-age children with family characteristics that put them at risk for reading difficulties (e.g., living in poverty, second-language English speakers, or immigrants) have increased by 50% to 75% (Allington, 2002; Dalaker, 2001; Wirt, Choy, Gerald, Provasnik, Rooney, Watanabe, & Tobin, 2002). Levels of absolute reading proficiency among adolescents have not been falling; however, literacy demands on adolescents have been increasing (Allington, 2002; Klenk & Kibby, 2000).

One source of increasing literacy demands on adolescent students is high-stakes testing (e.g., high school graduation examinations). Twenty-five states require high school students to pass tests that tap sophisticated reading skills in order to graduate (Center on Education Policy, 2004). The introduction of high school graduation exams is correlated with higher rates of dropping out among low-performing students (academic performance is the single best predictor of high school dropout; Battin-Pearson et al., 2000; Clarke, Haney, & Madaus, 2000; Reardon & Galindo, 2002).

A second source of demand for raising adolescents' reading skills is the No Child Left Behind Act (2001), which currently requires adequate yearly progress for all students in reading and math. The Act has put pressure on high schools to increase the literacy skills of all students. At the same time, retention in 9th grade has increased—the national retention rate was 4% in 1972-73 and 12% in 1997-98 (Haney et al., 2004). There is therefore an urgent need to develop and disseminate methods for increasing the literacy proficiency of young adolescents in the United States.

² These “trend scores” for NAEP are based on samples of 22,545 eighth-grade students in 1971, 22,693 students in 1984, and 5,933 students in 1999.

Need for Basic Research in Adolescent Literacy

Despite this 30-year history of documenting reading comprehension difficulties among young adolescents and the increasing literacy demands on them, most reading research has been conducted with children of elementary school age or younger, and has focused on word reading and other basic reading processes (see, e.g., the short literature reviews on adolescent literacy in Snow, 2002). The community of researchers in adolescent literacy has clearly identified this lack of basic research in reading comprehension in two recent documents: the RAND Reading for Understanding report (Snow, 2002) and a statement on research needs in adolescent literacy from the National Institute for Child Health and Human Development (NICHD, 2002). The RAND Reading for Understanding report, commissioned by the Office of Educational Research and Improvement (now the Institute of Education Sciences), points out that research is lacking on the relationship between comprehension and vocabulary knowledge, strategy use, motivation for reading, and how these develop over the adolescent years. Similarly, the NICHD statement on research needs, developed from workshops on adolescent literacy held in the spring of 2002 called for research “to understand the continued learning and development that takes place during adolescence in the areas of reading and writing.” (NICHD, 2002, p. 2)

Reading researchers have therefore clearly identified a need for basic research to understand reading comprehension in adolescents, specifically, research on different components of reading comprehension.

Prior Research on Components of Reading Comprehension in High School

Prior research on components of reading comprehension have considered both single variables (e.g., vocabulary), multiple variables (e.g., in regression studies), and

have used think-aloud methods. The early single-variable studies provided evidence (product data) for sets of variables that were later tested in multivariate models. Think-aloud studies, primarily working in a novice-expert paradigm, provided data about the processes involved in skilled reading. Few studies, however, have attempted to converge product (e.g., test score) and process (e.g., think-aloud) data in order to gain more information about participants' reading. Using multiple methods can provide convergent validity evidence for the two sources of information, and can thereby strengthen the conclusions about reading comprehension.

Single-variable studies. Most prior research in reading comprehension has investigated single variables (Pressley, 2000). At the high school level, the variables that have been investigated are background knowledge (e.g., Alexander & Kulikowich, 1991), inference (e.g., Davey, 1988), cognitive and metacognitive strategy use (e.g., Meyer, Brandt, & Bluth, 1980), vocabulary (e.g., Stahl, Hare, Sinatra, & Gregory, 1991), and word reading (e.g., Hood & Dubert, 1983). Each of these components has been the subject of a major research program in reading comprehension, has been found in experimental studies to contribute separately to young adolescents' reading comprehension, has led to the development of successful remediation programs which increase comprehension, and has led to findings that are consistent with those from both younger and older (e.g., undergraduate) students.³

³ Readers may wonder why motivation has not been included as a variable in this study. Motivation has also often been suggested as an important component in adolescent reading comprehension (e.g., Alvermann, Hinchman, Moore, Phelps, & Waff, 1998; Moje, 2000; Moje & O'Brien, 2001; Moore, Bean, Birdyshaw, & Rycik, 1999; Smith & Wilhelm, 2002), but there is a dearth of quantitative research. Most of the research that has been done on motivation and reading is correlational. The few experimental, longitudinal, and model-fitting studies that exist suggest the following: 1) the correlation between various aspects of motivation and reading increases as children get older (e.g., Paris & Oka, 1986), 2) children's level of various aspects of motivation for reading decreases over the school years (e.g., Wigfield, 1997), 3) self-efficacy is the most influential component of motivation (e.g., Shell, Murphy & Bruning, 1989), and 4) motivation does not have a direct effect on reading, but rather exerts its effects indirectly through strategy

Multi-component studies. Reading is also clearly more than any single component; several researchers have conducted multi-component studies of reading comprehension, though almost entirely with students of college age or older (see Artelt, Schiefele, & Schneider, 2001 for one high school study). One advantage of multi-component studies is that the variance unique to each predictor variable can be estimated (i.e., variance shared with other predictors, leading to deceptively high correlations, can be partialled out). These multi-component studies have also been mostly cross-sectional, which limits the inferences about causality that can be made from this body of evidence (Carr & Levy, 1990; but see Paris & Oka, 1986 for a 1-year longitudinal study with 3rd and 5th grade students). In addition, the majority of multi-component reading studies have tested the contribution of each component (e.g., using multiple regression), but have not tested the fit of the data to any particular model of reading comprehension.

Think-aloud studies. Multi-component studies can begin to show the complexity of reading comprehension. However, these studies rely on static measures of reading, which might fail to capture many on-line comprehension processes. This is especially true when closed-ended items are used (e.g., specific previewing and planning strategies that proficient readers use, but that are not asked about in questionnaire studies; see Hadwin, Winne, Stockley, Nesbit, & Woszczyna, 2001; Lundeberg, 1987; Winne, Jamieson-Noel, & Muis, 2002).

Another major approach to understanding reading comprehension is think-aloud, or process studies (Ericsson & Simon, 1993). Readers are asked to think aloud, or to verbalize what they are thinking, while reading a text. A comprehensive review of think-

use (e.g., van Kraayenoord & Schneider, 1999; Yopp & Dreher, 1994) or engagement (Guthrie & Wigfield, 2000).

aloud studies in reading by Pressley and Afflerbach (1995) showed the enormous range of strategic activities used by readers in elementary school (e.g., Langer, 1986), middle school (e.g., Loxterman, Beck, & McKeown, 1994), high school (e.g., Olshavsky, 1976-77), and among college students and adults (e.g., Afflerbach, 1990). Readers in think-aloud studies show evidence of planning their reading activities, enacting numerous cognitive and metacognitive strategies, monitoring the efficacy of those strategies, adjusting strategies flexibly, reflecting on and reacting to what was read, and many other processes (e.g., Azevedo, Cromley, & Seibert, 2004). Think-aloud studies have revealed reading processes of proficient readers that had not been identified by static measures, but they provide frequency data, which limits the statistical methods that may appropriately be used to analyze them. In addition, researchers have recently called for multi-method studies to triangulate data found in think-aloud protocols, questionnaire, and other measures (Graesser, Singer, & Trabasso, 1994; Long & Bourg, 1996; N. Perry, 2002; Whitney & Budd, 1996; Winne et al., 2002). When a study shows converging findings across multiple methods, this strengthens the conclusions of the study.

Coordinating product and process data. A small number of reading and cognitive psychology studies have coordinated *product* (e.g., standardized test) data with *process* (e.g., think-aloud) data in order to better understand and interpret the product data. For example, Novick and Holyoak (1991) measured college students' accuracy in solving analogous mathematics word problems, math SAT scores, and scores on a standardized analogy test (product data). They also collected think-aloud data and written notes (process data) that showed the extent to which students used analogical transfer in solving the problems. The product data by themselves showed a significant relationship of accuracy with SAT scores, but not with the analogy test. Think-aloud data, however,

showed that students who failed to transfer what they had learned from a source problem to target problems had specific difficulties in the *adaptation* phase of analogical transfer, even though they had similar scores on the analogy test. Thus, a single-method experiment may show *that* low component scores are related to low performance, but think-aloud data can provide evidence for *how* low component scores are related to low performance.

Previous reading research has rarely coordinated product and process data (but see Walczyk, Marsiglia, Bryan, & Naquin, 2001 for one example). However, this approach offers two distinct advantages. First, process data provide converging validity evidence for each product measure. Second, product and process data provide two views of the same activity (i.e., comprehension) and each data source can help us interpret the results obtained from analyzing the other.

The Inferential Mediation Model

In a previous study, Cromley and Azevedo (2004a) created a model of reading comprehension based on a literature review, termed the Inferential Mediation (IM) model (see Chapter III for a detailed discussion of this study). The IM model represents the interrelationships among the five variables (components of comprehension) listed on pp. 4-5: background knowledge, inference, strategy use, vocabulary, word reading accuracy, and reading comprehension (for operational definitions of the variables, see pp. 27-32). In that study, we had 63 9th-grade students at a range of reading comprehension levels complete researcher-developed measures of background knowledge and inference, and published measures of strategy use, vocabulary, word reading accuracy, and comprehension. Fourteen participants also provided think-aloud protocols. In that study, we then compared the fit of the IM model to two models of comprehension: Walter

Kintsch's (1988, 1998) Construction-Integration model and Charles Perfetti's (1985) Verbal Efficiency theory. The IM model was found to have a better fit to the data than did the CI model or VE theory.

The Construction-Integration Model and Verbal Efficiency Theory

With regard to existing theoretical models of reading comprehension, the IM model is closest to Kintsch's (1988, 1998) Construction-Integration (CI) Model and Perfetti's (1985) Verbal Efficiency (VE) Theory. All three models share the same set of predictors—background knowledge, inference, strategies, vocabulary, and word reading. In order to situate the IM model with regard to these other models, in the sections below I first outline the Construction-Integration model, the role each of the five predictor variables play in that model, and the experimental evidence supporting it. I then outline the same for Verbal Efficiency theory. Finally, I explain the two paths in the IM model that have weak or contradictory evidence, leading to the four variations on the IM model to be tested in the current study.

The Construction-Integration Model. Walter Kintsch's (1988, 1994, 1998; Kintsch et al., 1993) construction-integration (CI) model is a connectionist theory that proposes two phases in text comprehension: a construction phase and an integration phase. In the construction phase, reading a word (decoding) automatically activates that word and all of its meanings (vocabulary) in long-term memory (cf. Graesser, Millis, & Zwaan, 1997, p. 13). In addition, all of the semantic associates of the word (from background knowledge) are also activated. Semantic associations can be increased by teaching strategies that encourage readers to be active (see Kintsch's summarization training program, *Summary Street*; Wade-Stein & E. Kintsch, 2004; Questioning The Author [Beck, McKeown, Sandora, Kucan, & Worthy, 1996] is also suggested as a

specific strategy by Kintsch, 1998, on p. 329). From these three components—decoding, vocabulary, and background knowledge—a literal version of the text (the textbase) is constructed in the reader’s mind (cf. Kintsch, 1998, p. 127). Verbatim memory for text resides in the textbase (Kintsch, 1994).

As an example of the construction process, consider the sentence “Two masked gunmen made their getaway with \$100,000 from the First National Bank,” (Kintsch, 1998, p. 227). Two meanings of “bank” (financial institution and riverbank) are automatically and effortlessly activated from the reader’s long-term memory by reading the sentence. However, nodes cannot be activated if the reader does not have vocabulary knowledge, and fewer nodes will be activated if the reader has impoverished background knowledge (does not know that, e.g., rivers can flood and overflow their banks).

However, because *all* of a word’s meanings—including irrelevant meanings such as “riverbank” above—are activated in the construction phase, the mental representation is not yet coherent (cf. Kintsch, 1998, p. 103). In the second phase of comprehension, called integration, spreading activation among all components, together with inference processes, results in a stable activation pattern. Integration is a multi-cycle, slow, and sometimes effortful process. The resulting stable activation pattern, or coherent mental representation, is called a situation model. Comprehension, or understanding, resides in the situation model (Kintsch, 1994).

In the integration phase, background knowledge plays two roles: first, connections among nodes from long-term memory depend on background knowledge (see Whitney, Budd, Bramucci, & Crane, 1995). Second, background knowledge is used to draw inferences (elaborations) and to interpret the text (e.g., using domain-specific knowledge

about texts, Kintsch, 1994, p. 297; 1998, p. 167). Individual differences in background knowledge are therefore posited to have a strong impact on comprehension.

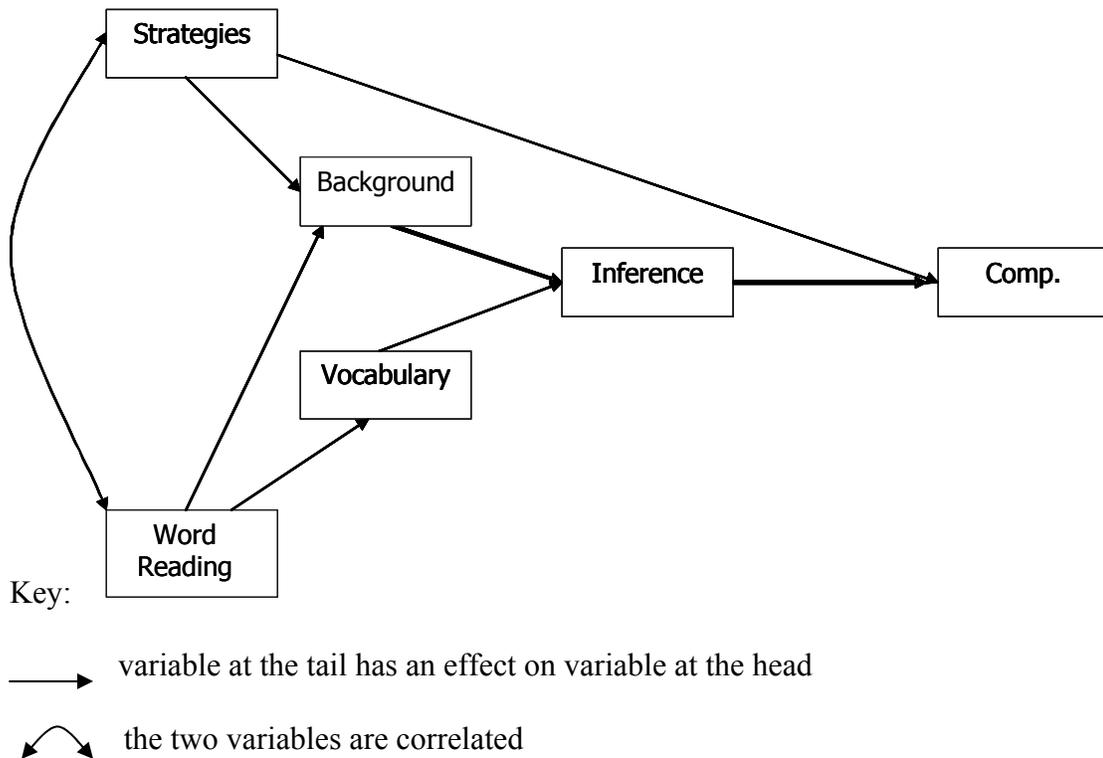
Domain-specific active problem-solving strategies may sometimes be used during the integration phase. “All text structures require domain-specific strategies and knowledge” (Kintsch, 1998, p. 167; e.g., expert strategies, Ericsson & Kintsch, 1995) For example, Kintsch specifically mentions that high-skill readers use imagery (1998, p. 108), domain-specific strategies for comprehending legal arguments (p. 191), domain-specific strategies to guide search processes (p. 191), activation of text schema and other strategies specific to understanding poetry (p. 213), and prediction (p. 244). Furthermore, when readers are not active, they can be taught strategies to make them active (p. 329). Kintsch mentions advance organizers (p. 321), summarizing (p. 277), Reciprocal Teaching (Brown & Palincsar, 1989; see Kintsch, 1998, p. 329), and Questioning The Author (Beck et al., 1996; see Kintsch, 1998, p. 329). Strategies are usually not necessary in familiar domains, however. For example, Kintsch argues that no special strategies are needed to make spatial inferences (1998, p. 214).

To return to the “bank” example, strong connections between the concepts “robbery,” “money,” and “bank” lead to a stable pattern of strong activation for the “financial institution” meaning of “bank,” and weak or zero activation for the “riverbank” meaning. This final, stable activation pattern is the *situation model*. After each sentence is read, a representation of it remains active in long-term working memory, and is combined with incoming information from the next sentence read (see Ericsson & Kintsch, 1995; Graesser et al., 1997).

Figure 1 represents the direct effects of the component processes in the CI model. Using the terminology of path analysis,⁴ each arrow in the diagram represents the theoretical statement that there is a direct effect (path) from the variable at the tail of the arrow to the variable at the head of the arrow. A curved, double-headed arrow represents the theoretical statement that two variables are correlated, but no reason is hypothesized for the correlation. Absent arrows between a pair of variables represent a theoretical statement that there is no direct effect (path) between them. Thus, Figure 1 represents Kintsch's theoretical stand that strategies have a direct effect on reading comprehension and an indirect effect on comprehension via background knowledge and then inference.

Figure 1

Path Diagram for the Construction-Integration Model



⁴ See pp. 23-26 for a detailed discussion of path diagrams.

By contrast, the diagram represents Kintsch's position that vocabulary has no direct effect on comprehension, but only an indirect effect via inference. The strongest emphasis in the CI model is on the role of background knowledge on inferences.

Evidence for the model. Many aspects of the CI model were tested on earlier versions of the theory (e.g., Kintsch & van Dijk, 1978). For example, there is ample evidence that readers can have an accurate textbase (as indicated by literal memory for text) but a poor situation model (as indicated by responses to comprehension questions; see Britton & Gülgöz, 1991; Wiley & Voss, 1999).

McNamara, Kintsch, Songer, and Kintsch (1996) specifically tested predictions from the CI model about the role of background knowledge in forming a textbase and situation model. The model predicts that a reader who is actively involved in forming a situation model (because the text does not deliver it fully formed) will be better able to answer inferential questions than a reader who reads passively (because the text is so explicit that it presents a fully-formed situation model). McNamara et al. measured the reading comprehension of students with different levels of background knowledge about the human circulatory system while reading text that was either more- or less-coherent and did or did not contain text signals referring to the text's macrostructure. High-knowledge students showed significantly better comprehension from texts that required effort because they were neither coherent nor contained text signals. Low-knowledge students, on the other hand, showed significantly better comprehension with coherent, signaled texts. For other experimental studies testing the CI model, see Britton and Gülgöz (1991), Caillies, Denhiere, and Kintsch (2002), Graesser, Kessler, Kreuz and McLain-Allen (1998), E. Kintsch (1990), Mannes and Kintsch (1987), McNamara (2001), Otero and Campanario (1990), and Singer and Halldorson (1996). For computer

simulation studies testing the CI model, see Otero and Kintsch (1992), Schmalhofer, McDaniel, and Keefe (2002), Singer and Halldorson (1996), and Singer and Kintsch (2001). Most of the evidence for the CI model is from undergraduate college students, but some studies were conducted with high school students.

Verbal Efficiency Theory.

Charles Perfetti's verbal efficiency theory (1985, 1988, 1989; Perfetti & Hart, 2001) is an interactive information-processing theory of reading comprehension.⁵ The essence of verbal efficiency theory is that when lexical access components are of high quality (i.e., rapid, simultaneous access to correct phonological [sound], orthographic [spelling], and semantic [meaning] representations of a word; Perfetti & Hart, 2001), then mental resources (called text work) are freed up for the higher-order components of comprehension, such as problem-solving strategies, inferences, and elaborations that depend on background knowledge.

Perfetti's claim is that lexical access is *the* driving force in reading acquisition (Perfetti, 1992) and *the* source of individual differences reading proficiency (Perfetti, 1989, 1994; Perfetti & Hart, 2001). Even though he argues that background knowledge "is a potent source of differences in reading ability" and that "Schemata are critical to comprehension," he goes on to argue that "It does not follow, however, that failures of schemata are a major explanatory mechanism for a general theory of reading ability"

⁵ The most detailed treatment of the theory was given in Perfetti's 1985 book, *Reading Ability*. Since that time, he has modified his treatment of the word reading and vocabulary (lexical quality; see Perfetti, 1992 and Perfetti & Hart, 2001) components of the theory, and has rarely discussed other "higher order" components of comprehension. However, in a book chapter written in 2001 (Perfetti & Hart, 2001), he affirms that he has not changed his interpretation of higher-order comprehension: "In its subsequent and current elaborations, efficiency of word identification . . . [is] about the quality of lexical representations . . . High-quality representations are . . . responsible for automaticity . . . which is what allows processing resources to be devoted to higher level comprehension" (p. 76). My discussion of verbal efficiency theory is therefore based predominantly on Perfetti's 1985 book.

(1985, p. 118). In short, Perfetti sees verbal efficiency (lexical quality) as the root of all reading problems, and inference, metacognitive strategies, and background knowledge as ineligible as sources of comprehension problems. Rather than verbal efficiency theory, therefore, this review focuses on Perfetti's components of reading comprehension (Perfetti, 1985, 1988). This a model that includes verbal efficiency theory, but also includes other components that Perfetti would not see as sources of reading comprehension difficulty.

In verbal efficiency theory (as in the CI model), reading proceeds in two phases. In the first phase, assembly, words on the page are recognized (requiring *word reading*) and matched with the mental lexicon, a process called *lexical access*. Words are then assembled into propositions with the help of *vocabulary* knowledge. Lexical access consists of decoding or word reading, followed by access to multiple word meanings (Perfetti & Hart, 2001). Perfetti uses the example of a story which begins, "Joe and his infant daughter were waiting for the doctor to get back from lunch" (Perfetti, 1985, p. 46). The reader must decode the words, access their meanings, and assemble the propositions. The resulting propositional representation is analogous to Kintsch's (1988) textbase (Perfetti, 1994). Assembly uses exclusively linguistic processes; Perfetti is emphatic that background knowledge has no role at all in lexical access (in Perfetti, 1989, 1992, he refers to the lexical access phase as "reading skill" or "general reading ability" as opposed to "comprehension"). Higher-skilled readers have higher quality lexical representations (e.g., more complete and accurate sound, spelling, and meaning knowledge) for more words than do lower-skilled readers (Perfetti, 1992; Perfetti & Hart, 2001). As a consequence, they are faster at creating propositional representations.

In the second phase, integration, the assembled propositions are integrated into a coherent whole (a *text model* which is analogous to Kintsch's [1988] situation model) through the process of inference-making. *Schema activation* (i.e., background knowledge) is required in order to draw these inferences. In the example above, father-daughter, doctor-visit, and doctor-lunch schemata are activated. Inferences are made (e.g., the visit was for medical care, rather than a social visit to a doctor friend), and an initial text model is constructed, including two characters and a complication (the doctor's absence). After decoding and encoding the propositions in the next sentence, "The room was warm and stuffy, so they opened the window" (Perfetti, 1988, p. 123), a second node is added to the existing text model, which now has two characters and two complications (the stuffy room and the absent doctor).

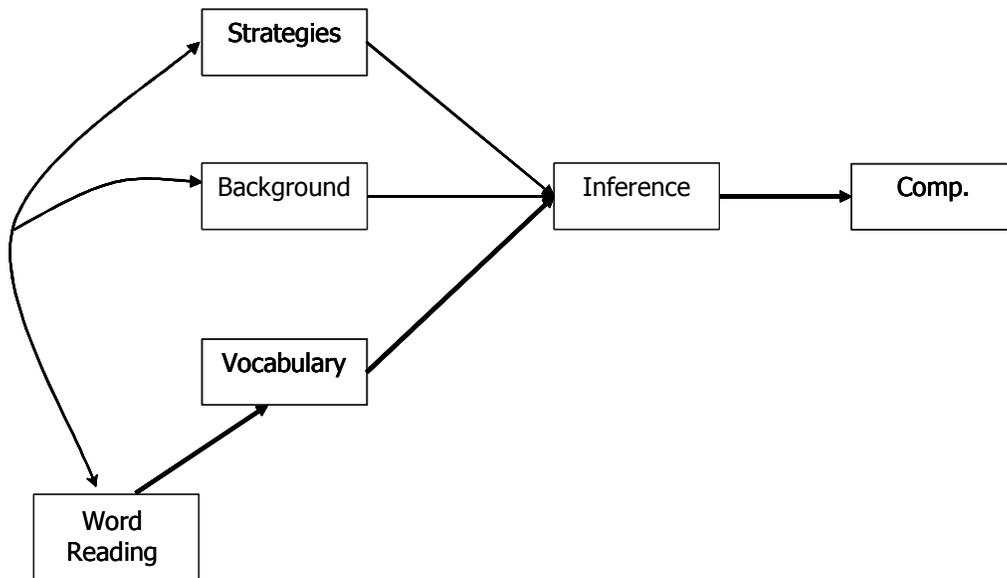
Part of (or perhaps after) forming a text model are the processes of "interpretive, inferential, and critical comprehension" (Perfetti, 1988, p. 122). These processes are more general cognitive (and less language-specific) processes, such as inference, knowledge, comprehension monitoring, and other problem solving strategies or "procedures to get the student reasoning about the content of a text" (Perfetti, 1994, p. 332). Perfetti also refers to reading comprehension strategies (e.g., metacognitive monitoring) as part of the reader's knowledge base, and therefore as having an influence on inference-making (Perfetti, 1985, p. 78-79).

As in the CI model, the text model (analogous to Kintsch's [1988] situation model) from the previous sentence remains active, and is modified when processing is complete for the next sentence. Also similar to the CI model both propositional encoding and comprehension processes take place in the context of a limited-capacity working memory (WM; Perfetti, 1985, 1994), however, WM is not *per se* a component of

comprehension. Figure 2 represents in a path diagram the direct effects among the component processes in Perfetti's components of reading comprehension model.

Figure 2

Path Diagram for the Verbal Efficiency-Based Model



Evidence for the theory. Perfetti (1994) has acknowledged that most support for verbal efficiency remains correlational. Most studies have tested the lexical access aspects of verbal efficiency—do higher-skilled readers have faster and more accurate access to phonological, orthographic, and semantic representations of words? (see Kuhara-Kojima, Hatano, Saito, & Haebara, 1996).

Bell and Perfetti (1994) examined reading comprehension components of 29 college students, 10 of whom scored high on both verbal and quantitative SAT scores (high-ability), 9 of whom scored low on both verbal and math (garden-variety poor readers), and 10 of whom scored low on verbal but normal on math (similar to dyslexics). Participants completed measures of reading comprehension, vocabulary (definitional and lexical access), background knowledge, letter matching, decoding (speed and accuracy),

memory span, and spelling. They also either read or listened to expository passages followed by either essay or multiple-choice questions. They conducted a series of regressions on different comprehension outcomes using listening comprehension, vocabulary, reading speed, and pseudoword decoding as predictors (the rationale for choosing these four predictors is not given). Depending on the outcome, these four predictors accounted for a significant 58% to 84% of variance (predictors were not significant for fiction comprehension). When entered last, listening comprehension accounted for a significant 11% to 37% of variance (depending on the outcome measure) and pseudoword decoding for a significant 11% of variance in science texts. Although this study provides some evidence for the continuing role of decoding in college student comprehension, it seems to provide more support for Perfetti's components of reading comprehension than for his verbal efficiency theory. However, there are serious problems (acknowledged by the authors) with using regression analysis with contrasting groups.

For additional tests of VE theory, see Haenggi and Perfetti (1994). For an expansion of verbal efficiency theory, see Jeffrey Walczyk's (1994, 2000; Walczyk & Taylor, 1996; Walczyk et al., 2001) research on his Compensatory-Encoding Model (C-EM). Most of the evidence for the Verbal Efficiency model is from undergraduate college students.

Differences Between the IM, CI, and VE Models

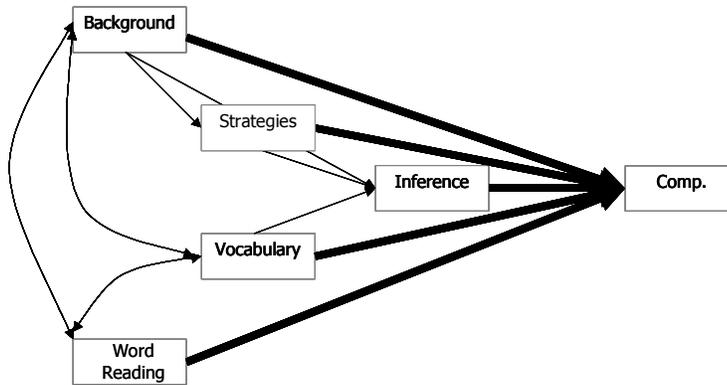
The primary difference between the IM model and the CI and VE models is that the IM model adds direct effects from each predictor to comprehension (see Figure 3; direct effects are indicated with **bold** lines). These are not part of the CI or VE models, except for the direct effect of strategies on comprehension in the CI model and the direct effect of inference in both models.

Including the direct effects in the IM model allows for readers to sometimes comprehend without having to make an inference; whereas according to the CI and VE models, inferences are always needed in order to comprehend. Including the direct effects is an advantage of the IM model, since readers do sometimes encounter simple declarative sentences which do not require inferences. For example, to comprehend the sentence, “The horses ate the grass,” readers must have background knowledge that horses ordinarily eat grass and vocabulary knowledge that horses are animals and grass is a plant. However, no inferences are needed in order to comprehend this simple declarative sentence. Note that some simple declarative sentences do require inferences, as in “The horses ate the cotton candy,” in which case the inference needs to be made that horses ordinarily eat food other than cotton candy, and that eating cotton candy is an unusual occurrence for a horse.

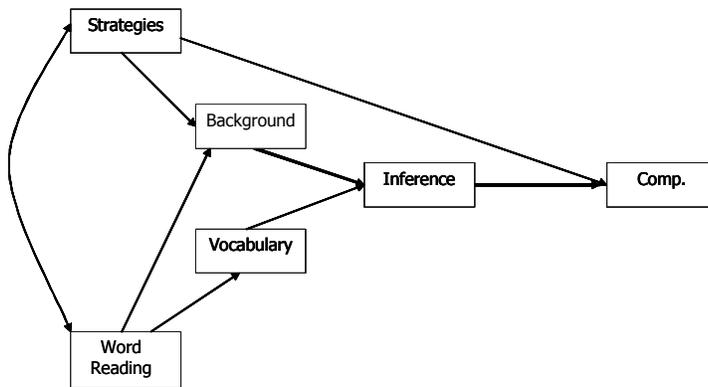
Figure 3

The Inferential Mediation Model, Construction-Integration Model, and Verbal Efficiency Models

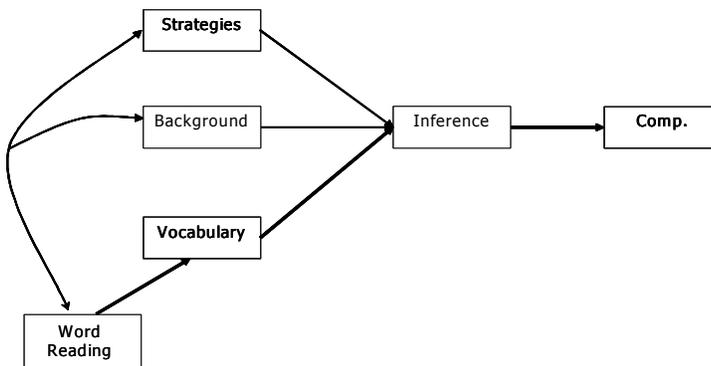
Inferential Mediation Model



Construction-Integration Model



Verbal Efficiency Model



Building on the Preliminary Study

The preliminary study showed that the IM model had a better fit to the data than did the CI or VE models. One goal of the dissertation study is to validate the IM model with a new sample drawn from the same population. A second goal was to then refine the model, since, there were two paths in the model that had weak or contradictory evidence: the effect of background knowledge on strategies, and the effect of word reading on vocabulary.

With regard to the first effect, there is evidence from 6 experimental studies that background knowledge has a direct effect on use of reading comprehension strategies. However, five experimental studies failed to find evidence for such an effect (see the literature review on pp. 47-52). Due to the conflicting evidence for this path, it was selected as one of the paths to be tested in the variations of the IM model.

With regard to the second effect, there is evidence from one experimental study that word reading has a direct effect on vocabulary (see the literature review on pp. 91-93). In addition, there is a strong theoretical expectation that word reading has a direct effect on vocabulary (Coltheart & Rastle, 1994; Perfetti, 1985). Finally, word reading has sometimes been found to have a direct effect on vocabulary with younger students (e.g., 2nd grade; Eldredge, Quinn, & Butterfield, 1990; but see Aarnoutse & Van Leeuwe, 2000). Due to the weak evidence for this path, it was selected as the second path to be tested in the variations of the IM model.

By testing the IM model with and without each of these two paths, there are four variations of the model to be tested: 1) Model 1, with both paths; 2) Model 2, with the effect of background knowledge on strategies but without the effect of word reading on

vocabulary, 3) Model 3, with the effect of word reading on vocabulary but without the effect of background knowledge on strategies, and 4) Model 4, with neither path.

The Present Study

Because of the need for studies of young adolescents' reading comprehension, and due to the limitations of single-method studies in reading, the present study uses a multi-component, multi-method approach to measure various components and academic reading comprehension—background knowledge, inference, strategy use, vocabulary, and word reading—for 9th grade students using both closed-ended measures and think-aloud protocols. First, the model comparison between the CI, VE, and IM models was replicated with a new sample. Then, four variations of the IM model were the fit to the data, and paths in the best-fitting model were investigated in the think-aloud protocols.

Research Questions

The research questions are:

1. Using a new sample, which has the best fit to the data: the CI, VE, or IM model?
2. What is the best-fitting of four related IM models for 9th grade readers?
3. What are the predictor variables that make the largest total contribution to reading comprehension in the best-fitting model for 9th grade readers?
4. How do high- and low-comprehending readers differ on those predictor variables?
5. How are those predictors revealed in the think-aloud protocols of 9th grade readers?

CHAPTER II: REVIEW OF THE LITERATURE

In order to design effective reading comprehension interventions for high school students, it is important to understand how different components affect comprehension, directly and indirectly. The purpose of this chapter is to review the theory and literature that supports the components selected for the current study—background knowledge, inference, strategies, vocabulary, and word reading—as well as the inter-relationships among the components; these inter-relationships comprise the Inferential Mediation (IM) model, a model of reading comprehension in academic text.

First, requirements for path analysis are explained, including the terminology and representations used in path diagrams. Next, the method for identifying the variables used to build the IM model is explained and the variables are defined. Then the IM model is presented, together with a summary table showing the number of studies supporting the model. Studies that provide evidence for each path within the IM model are then reviewed in detail and summarized, one path at a time.

Path Analysis

Path analysis is a statistical technique in which the fit of a model is tested using observed data. As in multiple linear regression, a set of predictor variables (exogenous variables) and a criterion variable (an endogenous variable) are identified. In addition to the direct effects that can be tested in regression, the effects of mediating variables can also be tested in path analysis (Kline, 1998; Pedhazur, 1997). Path analysis is a

confirmatory technique, in that each effect in the model must be based on prior theory and/or experimental research (Hair, Anderson, Tatham, & Black, 1998; Pedhazur, 1997; Wright, 1934). Sewall Wright, credited with originating path analysis, emphasized this in his 1934 summary of the method:

The method of path coefficients was developed primarily as a means of combining the quantitative information given by a system of correlation coefficients with such information as may be at hand with regard to the causal relations, and thus of making quantitative an interpretation which would otherwise be merely qualitative. (p. 175)

Note that path analysis uses a causal terminology (e.g., path, direct and indirect effects, influence, structural, uni-directional; see Hair et al., 1998; Kline, 1998; Pedhazur, 1997; Wright, 1934). However, modern path analysis researchers recognize that even a theoretically- and empirically-grounded path analysis can only meet some of the requirements for inferring causal relationships (see Kline, 1998); unidirectionality (the theoretical and research base must support a uni-directional effect; changing the direction of an arrow in the model changes the fit of the model) and specification of common causes (if it is hypothesized that X causes Y, but X and Y are actually affected by a common cause, Z, then the claim that X affects Y is untenable). Other requirements for causality, such as temporality (variables at the tail of an arrow occur before variables at the head of an arrow in time) are not met by cross-sectional path analysis studies.

Problems with common causes and temporality affect experimental studies as much as they do path analyses. Kline concludes, “It is only from a solid base of knowledge about theory and research that one can even begin to address these requirements” (Kline, 1998; p. 98).

Path Diagrams. Path analysts use a uniform system of symbols to represent models to be tested. In a path diagram, an observed variable is represented by a rectangle. An effect—the uni-directional influence¹ of one variable on another—is represented by a straight line with a one-headed arrow pointing *from* each exogenous variable *to* the variable it is hypothesized to affect. Correlations between variables are represented by curved, double-headed arrows.

Note that a uni-directional arrow in a path diagram represents an explicit theoretical statement that there is a direct effect (path) from the variable at the tail of the arrow to the variable at the head of the arrow. A unidirectional arrow also represents an explicit theoretical statement that there is *no* effect from the variable at the head of the arrow on the variable at the tail. An absent arrow between a pair of variables represents a theoretical statement that there is *no* direct effect (path) between them. A curved arrow represents an explicit theoretical statement that the variables are simply correlated. A curved arrow represents an explicit theoretical statement that neither variable is hypothesized to have an effect on the other (see Kline, 1998; Loehlin, 1998; Pedhazur, 1997).

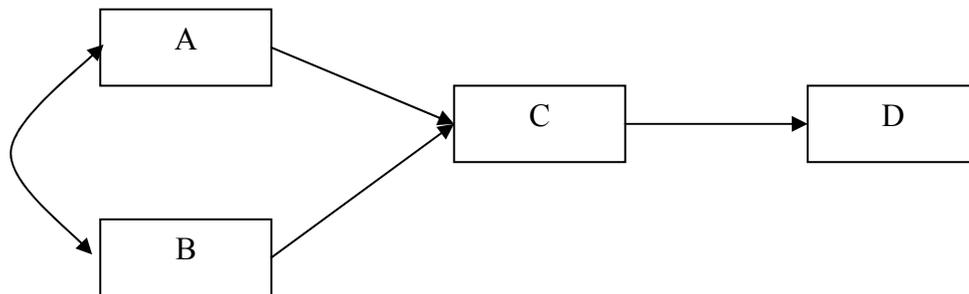
As an example, consider Figure 4. Variable A is hypothesized to have a direct effect on Variable C, and that Variable C is hypothesized to have no effect on Variable A. Variable B is also hypothesized to have a direct effect on Variable C, and Variable C is

¹ Note that educational psychologists are inconsistent in the terminology they use to verbally describe effects in path analysis. Across 6 path analysis or structural modeling studies in the *Journal of Educational Psychology* in 2004 (Bleeker & Jacobs, 2004; Chang et al, 2004; Levesque et al., 2004; Marsh & Hau, 2004; Urdan, 2004; Van Gelderen et al., 2004) 22 different terms were used to describe a uni-directional effect of one variable on another: affects (used in 2 articles), associated with (4), causes, consequences of, contribution of, determined by, effects (2), enhances, explains, fosters, impact of, importance of [X for Y], influences (3), leads to (3), makes (2), matters, mediates (2), predicts (2), produces, relation between (5), results from (2), and role of (2). Each article used an average of 7 different terms to describe uni-directional effects of one variable on another.

hypothesized to have no effect on Variable B. Variable C is hypothesized to have a direct effect on Variable D, and Variable D is hypothesized to have no effect on Variable C. Variables A and B are hypothesized to simply correlate with each other; Variable B is hypothesized to *not* have an effect on Variable A, and Variable A is hypothesized to *not* have an effect on Variable B. Correlations are termed *unanalyzed* relationships. Finally, Variables A and B are hypothesized to *not* have a direct effect on Variable D; however, they are hypothesized to have an *indirect* effect on D, an effect that is *mediated* by Variable C.

Figure 4

Illustrative Path Diagram



Key:

- variable at the tail has an effect on variable at the head
- ↔ the two variables are correlated

Building the Inferential Mediation (IM) Model.

The IM Model was built in two stages. In Stage 1, a set of predictor variables was identified, in which each variable has at least one experimental study showing it to affect high school students' reading comprehension. In Stage 2, experimental studies were sought that might support all possible inter-relationships among these predictor variables (10 pairs of variables x 2 directions) as well as the effects of the predictor variables on comprehension.

Stage 1: Identifying Predictor Variables for the IM Model. The first stage in building the IM Model was to identify variables that had been found to have an effect on high school students' reading comprehension. For each variable, studies had to show statistically significant results with high school students, and had to be published in a peer-reviewed journal. The variables identified were background knowledge (e.g., Schiefele, 1996), inference (e.g., Van den Broek, Tzeng, Risdien, Trabasso, & Basche, 2001), cognitive and metacognitive strategy use (e.g., Meyer et al., 1986), vocabulary (e.g., Graves, Boettcher, Peacock and Ryder, 1980), and word reading (e.g., Hood & Dubert, 1983). Some variables have been found to contribute to reading comprehension for older or younger students, but no studies were located that found these variables to be significant for high school students (these include phonemic awareness [e.g., Scarborough, Ehri, Olson, & Fowler, 1998] and working memory [e.g., Just & Carpenter, 1992]).

This set of five predictors—background knowledge, inference, strategy use, vocabulary, word reading—are also the variables found in two major theories of reading comprehension: Charles Perfetti's (1985) Verbal Efficiency theory and Walter Kintsch's (1988, 1998) Construction-Integration model (see pp. 9-18).

Definitions of the variables. Before presenting the evidence for the IM model, definitions of each variable will be given.

Background knowledge is all the world knowledge that the reader brings to the act of reading. It includes school-based knowledge and personal knowledge, episodic (events), declarative (facts) and procedural (how-to) knowledge (Alexander & Judy, 1988). Researchers have used one of two general approaches for investigating the relationship between background knowledge and reading comprehension—1) Is general

background knowledge (or world knowledge) related to general reading comprehension? (e.g., on an IQ test; Stanovich, West, & Harrison, 1995) or 2) Is background knowledge about a specific topic related to comprehension of a specific passage about that topic (e.g., Stevens, 1980)? Not surprisingly, fewer researchers have used the former approach, since developing acceptable, representative tests of general knowledge is a difficult task.

Inferencing is the logical process of combining information within sentences in text, between sentences in text, or between prior knowledge and text. For example, in order to understand who the word “he” is referring to in text, the reader must combine information in that sentence with information in a previous sentence that referred to a male. Readers also use inference processes to figure out the meaning of an unknown vocabulary word (Baumann, Edwards, Boland, Olejnik, & Kame'enui, 2003). Likewise, readers constantly add information from background knowledge to what they read in order to understand it. However, readers are often not aware of these processes. One important distinction made in the psychological literature is between on-line and off-line inferences (ones that are made only during later retrieval; Graesser et al., 1994). On-line inferences (like those made during concurrent think-aloud protocols) may include those made automatically as well as those made deliberately, strategically, and effortfully. Off-line inferences (like those made when answering post-reading questions or during retrospective protocols) are always seen as deliberate, strategic, and effortful. On-line inferences are the ones of highest theoretical interest to psychologists (e.g., Long, Seely, & Oppy, 1999), presumably because they represent the situation in most reading contexts.

Off-line inferences have been of great interest to educational psychologists, since school-related reading often includes reading passages followed by questions that require

inferences (e.g., Hare, Rabinowitz, & Schieble, 1989). In the literature reviewed below many studies considered both on-line and off-line inferences, and the measures proposed for the study (think-aloud and component measures) are designed to capture both types of inference. For a review from a psychological perspective, see Graesser et al. (1994). For reviews from an educational psychology perspective, see Mayer (1998), Oakhill and Yuill (1996), and Stothard (1994).

Strategies. Proficient readers use cognitive and metacognitive *reading strategies* such as setting goals before they begin to read, asking themselves questions and answering them while reading, summarizing, and reflecting on what they read. Alexander and Judy (1988) define strategies as “goal-directed procedures that are planfully or intentionally evoked . . . [that] aid in the regulation, execution, or evaluation of [a] task” (p. 376). Strategies help proficient readers understand better what they read. Readers are not necessarily aware of using these strategies, although they are able to verbalize many of them when asked to think aloud during reading, and are able to identify some of them on questionnaires. Some research suggests, however, that while good readers can accurately self-report strategies, poor readers have weaknesses in metacognitive processes that lead them to inaccurately self-report strategy use (Baker & Cerro, 2000). Strategies can be taught to children who struggle with comprehension, which improves their understanding of texts.

Vocabulary is often defined as knowledge of a word’s meaning. However, there are many aspects of word knowledge, most of which have received little attention from researchers. Nagy and Scott (2000) point out that knowledge about any single word is multidimensional (e.g., giving a definition, knowing the part of speech, being able to use the word correctly), incremental (not all-or-nothing), polysemous (many

words have more than one meaning), interrelated (e.g., understanding a definition requires understanding other words in the definition), and heterogeneous (e.g., the knowledge one can have about function words, technical terms, and concrete nouns varies). In addition to word knowledge, vocabulary knowledge includes knowing the meanings of affixes (prefixes and suffixes), understanding relationships between words (e.g., democracy and democratic, called morphological knowledge), and strategies for figuring out new words (Nagy, Diakidoy, & Anderson, 1993). The majority of research on vocabulary and reading comprehension focuses on single meanings of words. For reviews of vocabulary and comprehension, see Nagy and Scott (2000), National Reading Panel [NRP] (2000), and Stahl (1998).

Word reading includes both a reader's sight words (stored in long-term memory) and word attack skills. The latter include decoding, analogy, and morphological strategies (e.g., using prefixes and suffixes; Nagy et al., 1993). Measures of word reading often include real words and nonsense words or pseudowords (e.g., blum or grame) that follow regular spelling-sound patterns in English. Especially with older students, nonsense words are thought to reveal students' true word attack skills, since any real word could already be a sight word for the student (Shankweiler, Lundquist, Dreyer, & Dickinson, 1996). Word reading is distinct from vocabulary knowledge in that a reader may be able to read a word but not know its meaning, or may know the meaning of a word if it is spoken out loud but may not be able to pronounce the word in its written form. For reviews of word reading and comprehension see Blachman (2000), NRP (2000), and Pressley (2000).

Reading Comprehension. Defining *reading comprehension* has been a contentious process. Discourse processing researchers almost unanimously define

comprehension as the formation of an internally consistent mental representation of text, through the process that combines information from text with the reader's prior knowledge. For example, the RAND Reading Research Group used the following definition:

The process of simultaneously extracting and constructing meaning through interaction and involvement with written language. We use the words *extracting* and *constructing* to emphasize both the importance and the insufficiency of the text as a determinant of reading comprehension.

Comprehension entails three elements:

- The *reader* who is doing the comprehending
- The *text* that is to be comprehended
- The *activity* in which comprehension is a part.

In considering the reader, we include all the capacities, abilities, knowledge, and experiences that a person brings to the act of reading. Text is broadly construed to include any printed text or electronic text. In considering activity, we include the purposes, processes, and consequences associated with the act of reading. (Snow, 2002, p. 11)

(See Gernsbacher, Varner, & Faust, 1990; Graesser & Britton, 1996; Graesser et al., 1997; Kintsch et al., & Nathan, 1993; Konold, Juel, McKinnon, & Deffes, 2003; van den Broek, Ridsen, Fletcher, & Thurlow, 1995; and Zwaan & Brown, 1996 for highly similar definitions). However, as Harris and Hodges (1995) point out, researchers are split between those who feel a comprehender's mental representation must match that intended by the author (the perspective adopted in the current study), and those from a more literary bent (e.g., Lee, 2001) who feel that any internally consistent representation is a sign of comprehension. The majority of psychological researchers cited in this

literature review adopt the former definition, explicitly or implicitly.

An important issue in comprehension research is the measurement instrument—the RAND Reading Research Group (Snow, 2002) argued forcefully that there are no existing measures of reading comprehension that are firmly grounded in theories of reading comprehension (but see Sheehan & Mislevy, 1990 for a theory-based example). By default, most researchers use nationally normed, standardized measures (e.g., Nelson-Denny, Gates-MacGinitie), researcher-developed comprehension questions (either multiple-choice or open-ended), or oral or free recall or summarization tasks to measure comprehension.

Having defined the variables, I will now present the IM Model and summarize the empirical support for the model. For each path, I first review experimental studies that establish each path and then review path analytic, regression, think -aloud, and correlational studies that provide corroboration for that path.

Stage 2: Evidence for the IM Model. Below, I review the criteria for selecting studies, the research evidence for each path, and the rationale for the four variations of the model to be tested in the current study.

Criteria for selecting studies. The IM model is based on a comprehensive review of quantitative studies published in peer-reviewed journals, and also includes book chapters reporting multi-component reading comprehension studies. Searches were conducted in the ERIC and PsycINFO databases through April, 2002, with no limitations on the time period in which studies could be published.

Studies were restricted to those with participants in grades 4 and above, since the effect of word reading on the other four components is expected to be quite different for students in the “learning to read” phase than in the “reading to learn” phase. Studies were

also restricted to non-disabled, non-second language participants (e.g., pull-out English for Speakers of Other Languages classes in the US or English language learners in other countries).

To build the model, studies of seven types were sought. First, for each pair of variables, experimental or quasi-experimental studies using inferential statistics had to be located to provide evidence of an effect for one variable on another. In no case was there evidence for any bi-directional effects among any pair of predictors. After establishing that there was experimental evidence for an effect, I then identified corroborating studies, including multi-component reading comprehension studies (regression, path analysis, and Structural Equation Modeling studies), think-aloud studies, and finally studies that were simply correlational.

For two effects—the effect of background knowledge on strategies and the effect of word reading on vocabulary—findings were either contradictory or weak (see the reviews of specific studies below). These two paths were therefore specifically selected to be tested in the present study.

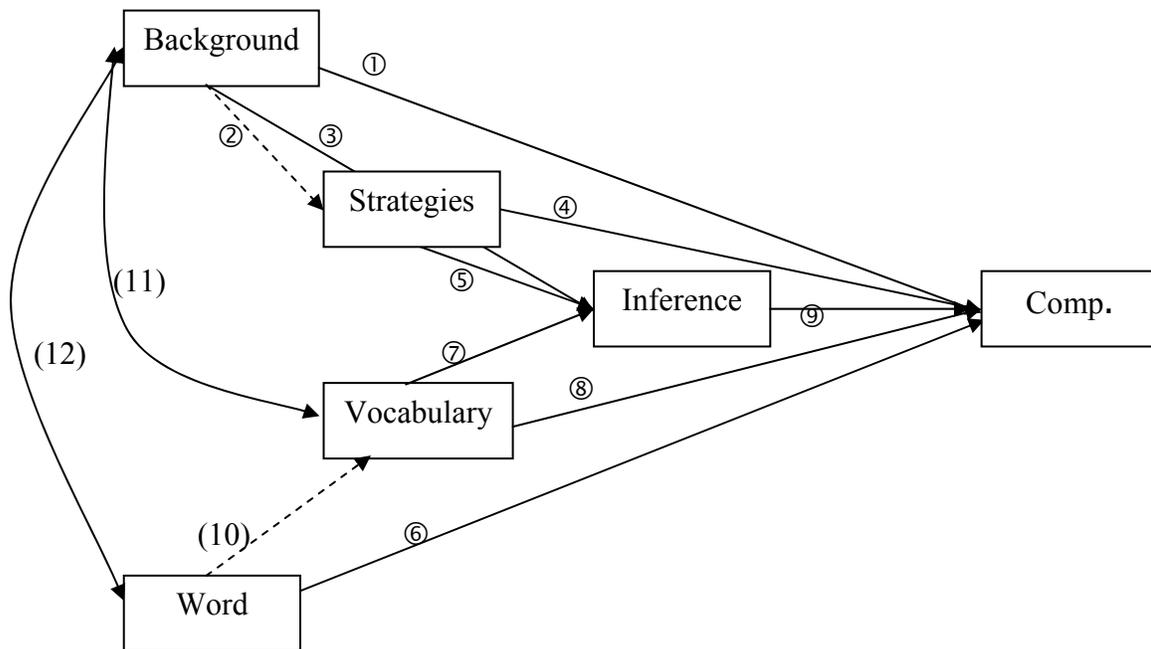
The IM Model. A path diagram representing the IM model is presented in Figure 5. Each path in the model is numbered to facilitate discussion of studies supporting the path. Table 1 summarizes the number of experimental studies that establish each path and the number of model fitting, regression, think-aloud, and correlational studies that provide corroboration. The fit of the model was initially tested in a pilot study conducted with 63 9th-grade students in spring, 2003 (Cromley & Azevedo, 2004a).

In the diagram, the presence of a uni-directional arrow indicates that at least one experimental study was identified, in the direction indicated by the arrow, for grades 4 or

higher. The absence of an arrow indicates that either no experimental studies were identified, or only studies in 3rd grade or below were identified. For example, the arrow from vocabulary to inference indicates that evidence was found for a direct effect of vocabulary on inference. The diagram also indicates that no evidence was found for a direct effect of inference on vocabulary, nor for a direct effect of vocabulary on strategies or background knowledge.

Figure 5

Path Diagram for the Inferential Mediation Model (Dashed Arrows Indicate Paths to Be Tested)



For the exogenous variables (i.e., background knowledge, vocabulary and word reading), once it was clear that there was not evidence for any of these variables having an effect on each other, correlational studies were sought to support the correlations in the model (indicated by curved arrows).

Table 1

Number of Experimental Studies Establishing Each Path and Corroborating Studies

	<i>Exper. non- inter- vention</i>	<i>Exper. inter- vention</i>	<i>Total experi- mental</i>	<i>Path model/ SEM</i>	<i>Regress -ion</i>	<i>Think- aloud</i>	<i>Correla- tional</i>
Path 1: Effect of background knowledge on comprehension							
Total	11	11	22	1	3	7	6
HS	3	7	10	0	0	5	4
Path 2: Effect of background knowledge on strategies							
Total	5	1	6	0	2	0	3
HS	0	0	0	0	1	0	2
Path 3: Effect of background knowledge on inference							
Total	5	3	8	0	2	0	1
HS	1	1	2	0	0	0	0
Path 4: Effect of strategies on comprehension							
Total	5	46	50	3	3	13	8
HS	2	25	27	2	0	9	6
Path 5: Effect of strategies on inference							
Total	1	5	6	1	0	0	2
HS	0	1	1	0	0	0	0
Path 6: Effect of word reading on comprehension							
Total	6	3	9	1	10	1	6
HS	1	0	1	1	2	1	2
Path 7: Effect of vocabulary on inference							
Total	1	3	4	0	0	0	4
HS	0	0	0	0	0	0	1
Path 8: Effect of vocabulary on comprehension							
Total	7	5	12	1	5	6	13
HS	3	0	3	1	1	6	8
Path 9: Effect of inference on comprehension							
Total	7	5	12	1	3	9	6
HS	3	2	5	0	0	6	0
Path 10: Effect of word reading on vocabulary							
Total	1	0	1	0	1	0	2
HS	0	0	0	0	1	0	0
Total of effects							
Total	48	82	130	8	29	35	51
HS	13	36	49	4	5	27	23
Path 11: Correlation between background knowledge and vocabulary							
Total							3
HS							1
Path 12: Correlation between background knowledge and word reading							
Total							2
HS							0

Note: HS = Studies with High School students.

Path 1: Effect of background knowledge on comprehension. The role of background knowledge in reading comprehension has been studied since the early days of the “cognitive revolution” (e.g., Anderson, Reynolds, Schallert, & Goetz, 1977; Bransford & Johnson, 1973). In those early studies, college students given artificial, intentionally ambiguous texts (e.g., “Doing Laundry”) remembered significantly more when given a title that activated relevant background knowledge.

Experimental non-intervention studies. Several experimental studies with young adolescent students demonstrate the significant effect of background knowledge on reading comprehension. Recht and Leslie (1988) compared 64 7th and 8th grade students who were high and low in baseball knowledge and varied in reading ability (above the 70th percentile or below the 30th percentile on the SRA comprehension subtest). Participants read a 625-word passage describing a baseball game, then reenacted the game using a model and described what was happening, verbally summarized the passage, and sorted 22 sentences according to their importance. Significant main effects of prior knowledge were found on all measures; low-ability students with higher prior knowledge outscored high-ability students with low prior knowledge on all measures.

Adams, Bell and Perfetti (1995) conducted a study with a similar design with 106 boys (24 4th grade, 29 5th grade, 22 6th grade, and 26 7th grade), measuring prior football knowledge and reading football-related and non-football-related texts. For the football text, there was a significant main effect of background knowledge on answers to comprehension questions. Participants with low background knowledge but high comprehension scored as well as those with high background knowledge but low comprehension, but only on the football text. Similarly, Schneider, Körkel and Weinert (1989; see below, p. 47) in Experiment 1 found that students with more background

knowledge about soccer were significantly better able to recall text details from a soccer text, regardless of reading skill. In Experiment 2 those with more background knowledge were better able to recall text details and performed better on a cloze test, regardless of reading skill. Adams et al. and Schneider et al. both argue that high background knowledge can somewhat compensate for low verbal aptitude.

Callahan and Drum (1984) compared 10 high-ability ($\geq 90^{\text{th}}$ percentile on MAT) to 10 average-ability (30^{th} - 60^{th} percentile) 5^{th} and 6^{th} grade students reading social studies textbook passages. There was a significant difference between high- and average ability students in prior knowledge. Prior knowledge was the only significant predictor to enter a regression with free recall as the criterion ($R^2 = .52$).

Freebody and Anderson (1983a) in Experiment 2 ($N = 88$) gave above-average 6^{th} grade students social studies passages written about familiar and unfamiliar settings (e.g., a supermarket or a musical ceremony in an African village). There was a significant main effect of familiarity on recall and sentence verification. Curiously, students who read about unfamiliar settings wrote better summaries, but this might be explained by Kintsch's (1988, 1998) Construction-Integration model—students had to engage in deeper processing in order to understand the passages well enough to write a 3-sentence summary of them.

Stevens (1980) constructed a 100-item background knowledge test about 25 specific topics, and examined its relationship to reading comprehension questions from standardized grade-level reading passages about those topics. All 108 9^{th} grade students from a small public high school were grouped into low, medium, and high comprehension groups using the Nelson-Denny reading test (no criteria are given for the divisions). After completing the background knowledge test, students were assigned two

specific passages, one about which they had high knowledge and one about which they had low knowledge (that is, not all students read the same passages). There were significant main effects for both prior knowledge and ability, but no interaction.

Alexander and Kulikowich (1991) explored the effect of background knowledge, analogy, and reading comprehension on biology texts with 75 9th and 10th grade biology students. Background knowledge was measured with a 25-item, researcher-developed multiple-choice biology definitions test. Reading comprehension was assessed with two passages, each followed by 17 multiple-choice questions. Only analogical skill significantly discriminated between above- and below-average comprehension scores. In a similar experiment with 6th grade students, prior knowledge did significantly discriminate; with college students it did not. The authors conclude that there is a developmental trend for the relationship between knowledge/strategies and comprehension. Schiefele (1996) found significant effects of prior knowledge on sentence recognition (a measure of textbase) for 107 twelfth-grade German students reading a text about television, but no such effect for a text about dinosaurs. There was no significant effect of background knowledge on verification (a measure of situation model) for either text.

Similar results have been found for college students. McNamara (2001) found a significant main effect of prior knowledge on reading comprehension as measured by short-answer comprehension questions about cell division with 80 undergraduate students. Caillies et al. (2002) found a significant main effect of prior knowledge about computers on comprehension for 54 college students. Likewise, Kuhara-Kojima and Hatano (1991) found a significant main effect for prior baseball knowledge on 158

college students' ability to match pairs of facts from the text after reading a baseball passage.

Experimental intervention studies. A number of studies have found that teaching students the strategy of activating prior knowledge is effective in increasing reading comprehension. They are summarized in Table 2 below.

Table 2

Prior Knowledge Strategy Instruction Interventions Showing Effects on Reading Comprehension, Ranked by Age

Authors	Participants	Conditions, materials, posttest(s)	Results
Dole, Brown, & Trathen, 1996	57 5 th & 6 th grade from at-risk school	Pre-teach vs. Multiple strategies (predict, main idea & graphic organizer) vs. Control (basal); Basal stories; Short answer immediate & delayed	SI > PT = C on immediate and delayed
Dole, Valencia, Greer, & Wardrop, 1991	63 5 th grade from 3 average classes (Stanford Achievement Test)	Pre-teach vs. Prior knowledge activation vs. Control; Varied (basal); MC comprehension	PT > PKA > C
Stahl, Jacobson, Davis, & Davis, 1989	Exp. 1: 90 6 th grade above average (Gates-MacGinitie)	Pre-teach relevant v. Pre-teach irrelevant; Social studies text; Multiple-choice, sentence verification, recall	Pre-teach relevant group signif. better on recall only

Authors	Participants	Conditions, materials, posttest(s)	Results
Graves, Cook, & LaBerge, 1983	Experiment 1: 32 8 th grade low Experiment 2: 40 8 th grade very low (on SRA)	Prior knowledge activation (PKA) vs. Control; Short stories; Experiment 1: Multiple-choice comprehension Experiment 2: Free recall, short answer	T > C on all measures
Schmidt, De Volder, De Grave, Moust, & Patel, 1989	88 9 th (<i>n</i> = 46) & 10 th (<i>n</i> = 42)	PKA Self-explanation (elaboration); Biology (RBC's & osmosis) Free recall	Strategy group had signif. more explanations in recall
Slater, Graves, & Piche, 1985	224 9 th low, middle, hi on California Ach Test 89, 69 & 46 %iles	PKA; Social studies (California gold rush); 20-item multiple choice posttest & written recall	Strategy group signif. better on mult. choice, amount recalled & hi level recall
Spires & Donley, 1998	St 1: 79 St 2: 161 St 1: 9 th Equal numbers of hi, mid, lo per California Ach Test St 2: Same	PKA Main idea (St 1) Add MI+PKA (St 2); St 1: Social studies (family ecology, equal rights) St 2: Same; No access to texts Immediate & 4 week delayed Social Studies: multiple choice literal & open ended Short story: mult choice literal & open ended application questions St 2: Same	St 1: All literal—Main idea sig > PKA = ctrl. All application—PKA sig > Main idea = ctrl. St 2: All literal—all treatments sig > control Application—MIPKA mostly = PKA mostly > MI mostly > control Biggest results on delayed tests

Authors	Participants	Conditions, materials, posttest(s)	Results
Roberts, 1988	117 9 th grade college prep. Teachers reported no low readers	Pre-teach vs. PKA; Sports (soccer); 13 short answer fact & inference questions	Pre-teach group signif. better on inference only
Hayes & Tierney, 1982	100 11 th -12 th average and above CTBS	PKA; Sports (cricket); 50-item prediction measure (multiple choice) and recall	Strategy group signif. better on both measures
Lee, 1995	77 12 th grade inner-city African-American	PKA; Literature; 8-item short-answer literature interpretation post-test	Strategy group signif. better
Bean, Searles, Singer, & Cowen, 1990	111 HS (no grade or age spec.)	PKA (with lecture, analogies and pictures); Biology (parts of cell); 7-item matching test + 1 essay (no pretest)	$L + A + P > L+A = A > \text{control}$

Path analysis studies. Britton, Stimson, Stennet, and Gülgöz

(1998) used structural equation modeling (SEM) to test a model of reading comprehension with 211 Air Force recruits age 17-25 who read about the Vietnam War. Predictor variables included metacognitive activities, inference, domain knowledge, and working memory. Domain knowledge loaded a significant .66 on text learning. The model had a slightly less-than-ideal fit: CFI = .94, RMSEA = .05.

Regression studies. A regression study is tantamount to testing a path model in which each predictor is hypothesized to have only a direct effect on the

criterion variable, while not allowing for any indirect effects (Pedhazur, 1997). Note further that regression is not a symmetrical procedure; reversing the analysis and using the criterion variable as a predictor would result in a different solution. For studies with explanatory aims, Pedhazur recommends only partitioning of variance as a methodological approach. Some of the regression studies reviewed below, therefore, may provide evidence of a direct effect (those which enter the variable of interest last), whereas others may provide no more than correlational evidence.

Haenggi and Perfetti (1994) compared 34 undergraduates who tested above ($n = 17$) or below ($n = 17$) the median on the Nelson-Denny reading comprehension test. Participants completed word and pseudoword reading tests, and measures of semantic judgment, sentence verification, a researcher-developed, text-specific background knowledge measure, and a test of short-term retention of text. They also read a long (3,265-word) history text on the construction of the Panama Canal, and answered 20 multiple-choice questions about it. Prior knowledge and memory for recently read text (the two variables with significant correlations) were regressed on comprehension, accounting for a significant 71% of variance.

However, Bell and Perfetti (1994) did not find that background knowledge made a significant contribution to reading comprehension for 29 college students. Ten of the students scored high on both verbal and quantitative SAT scores (high-ability), 9 scored low on both verbal and math (garden-variety poor readers), and 10 scored low on verbal but normal on math (similar to dyslexics). Participants completed the Nelson-Denny reading comprehension and vocabulary tests; a researcher-developed, text-specific background knowledge measure; letter matching; decoding (speed and accuracy for words and pseudowords); lexical decision, memory span, and spelling measures. They

also either read or listened to ten 200-word expository passages followed by essay questions and six 2,000 word expository passages followed by multiple-choice questions. They conducted a series of regressions on different comprehension outcomes using listening comprehension, vocabulary, reading speed, and pseudoword decoding as predictors (the rationale for choosing these four predictors is not given). Depending on the outcome, these four predictors accounted for a significant 58.2% to 84.0% of variance (predictors were not significant for fiction comprehension). When entered last, listening comprehension accounted for a significant 10.9% to 36.9% of variance (depending on the outcome measure) and pseudoword decoding for a significant 10.6% of variance in science texts.

Peverly, Brobst, Graham and Shaw (2003) regressed text-specific background knowledge, study time, and student confidence ratings on factual knowledge scores from a free recall and multiple-choice measure given to 82 undergraduate students. Background knowledge did not explain a significant amount of variance for either students who were allowed to take notes while learning, or for students who were in a no-notes condition ($n = 41$ in each group), and R^2 was a low .14 and .29 for the two groups, respectively.

Think-aloud studies. Several think-aloud studies have also found differences between good and poor readers on use of background knowledge. Note that there are methodological problems with all of these studies that preclude reporting statistically significant differences—either inferential statistics were calculated for raw frequency data, or utterances (rather than participants) were used as the unit of analysis. Peskin (1998) used a verbal protocol methodology to study 8 experts (2nd year English graduate students) and 8 novices (11th-12th grade high school students) as they read 2

poems. Like experts in many other domains (e.g., Ericsson, 1996; Ericsson & Smith, 1991), poetry experts had not only more domain knowledge of poetry (e.g., names of poems and poets, meter and rhyme schemes, types of poetry), but their knowledge was well-integrated and easily accessed, which enabled them to better comprehend the poems that did novices.

Afflerbach (1990) had 15 skilled readers—10 graduate students and 5 high school students from a gifted and talented program—think out loud and make predictions while reading three essays and two short stories. Participants were asked to rate the familiarity of the content as a measure of prior knowledge. All readers showed the same pattern—passages with the highest familiarity were associated with more predictions; there were no effects for type of passage. In the think-aloud protocols, these skilled readers verbalized their background knowledge and also monitored their level of background knowledge (e.g., “This isn’t familiar at all . . . the story doesn’t make sense” p. 143). Mosborg (2002) found similar results with ten gifted high school students thinking aloud while reading newspaper articles about current events.

Using 30 gifted (95th percentile on ITBS comprehension) and 30 average (40th-60th percentile on ITBS) readers in grades 8, 10 and 12, Fehrenbach (1991) found that readers used a strategy of “relating to content area,” defined as “add information related to text based on content area knowledge or personal knowledge” (p. 126). “Relating to content area” was used more by gifted than by average students.

Kletzien (1991, 1992) studied activation of prior knowledge by good and poor 10th-11th grade comprehenders as they read social studies texts of varying difficulty. Reading groups were defined by the California Test of Basic Skills (above 75th percentile; below 50th percentile). In Kletzien (1991) there were no differences in

activation of prior knowledge between good and poor readers on independent or instructional level texts. On frustration level texts, poor readers used less prior knowledge activation than did good readers. In another study, Kletzien (1992) found that good readers used more prior knowledge activation only on texts having a “collection” structure, not those with a “causation” structure.

Earthman (1992) studied 8 college freshmen and 8 literature graduate students thinking aloud while reading poetry. The graduate students more often activated background knowledge about both the topics of the poems (e.g., the Jewish Sabbath) and about poetry.

Correlational studies. Significant correlations between background knowledge and reading comprehension have been frequently reported in the literature and are summarized in Table 3. Correlations in absence of theory are not particularly meaningful; however, they are presented here as corroboration in the context of the effect of background knowledge and reading comprehension supported by the experimental studies reviewed above. In the correlations, there is a suggestion of a developmental trend of increasing correlation with age. This could be due to a direct effect, an indirect effect (i.e., the effect of background knowledge is mediated by some other variable[s]), or due to a common third factor.

Table 3

*Reported Correlations between Background Knowledge and Reading Comprehension,
Ranked by Age*

Study	Correlation	Participants	Text
Langer, 1984	.34 WWI passage .41 Stonehenge passage	124 6 th grade students	Two passages: WWI and Stonehenge
Leslie & Cooper, 1993	.57	72 9 th grade students	Fiction, biology, and history passages
Kozminsky & Kozminsky, 2001	.69 across groups .55 academic students .57 semi-academic students .17 (ns) vocational students -.07 LD (ns) students	205 9 th grade 33 academic 121 semi-academic 21 vocational 22 LD	Two existing Hebrew multiple-choice reading comprehension measures.
Cunningham & Stanovich, 1997	.69-.75, depending on background measure	27 11 th grade students	Nelson-Denny reading comprehension (narrative only)
Langer, 1980	.44 parakeets passage .74 schizophrenia passage	36 high school seniors in Advanced Placement English classes	Two passages: parakeets and schizophrenia
Graesser & Bertus, 1998	.43 for younger adults .42 for older adults	40 younger adults (<i>M</i> age = 22) 40 older adults (<i>M</i> age = 67)	24 five-sentence science passages

In summary, there is evidence from 22 experimental studies that prior knowledge has a direct effect on reading comprehension. Seventeen path analysis, regression, think-aloud and correlational studies were also found that were consistent with this effect.

Significant effects have been found for both text-specific and general background knowledge measures across studies with students in 6th grade through older adults. They were found for a variety of measures of comprehension, including multiple-choice tests, verbal and written free recall, card sorting, and other tasks.

Path 2: Effect of background knowledge on strategies. Background knowledge might be thought to have an effect on the use of reading comprehension strategies for the same reasons it is thought to have an effect on comprehension. For example, in order to apply the strategy of summarizing, the reader must identify important information. But without any domain knowledge, important information and extraneous detail are hard to separate.

Experimental non-intervention studies. Schneider et al. (1989) compared 3rd, 5th, and 7th grade German students who were either low or high in knowledge about soccer and low or high in verbal aptitude (including tests of vocabulary and comprehension). In experiment 1, students (106 in 3rd grade, 236 in 5th grade, and 234 in 7th grade) simultaneously listened to and read a story about a soccer game with 3 embedded contradictions; the dependent variable was number of contradictions detected, which requires the strategy of metacognitive monitoring. There was a significant main effect for background knowledge on detecting contradictions and no significant main effect for aptitude. That is, high-knowledge but low skill readers detected as many contradictions as did high-knowledge high-skilled readers. In Experiment 2, there were 64 3rd-grade, 67 5th-grade, and 54 7th-grade students, who also rated the importance of sentences in the text (a main idea strategy). There was a significant main effect for background knowledge on importance rating and no significant main effect for aptitude.

That is, high-knowledge but low skill readers rated the importance of sentences as well as did high-knowledge high-skilled readers.

Symons and Pressley (1993) compared the effectiveness of search strategies used by undergraduates high and low in prior knowledge. Participants were students in an educational psychology course (fall participants, $n = 36$, winter participants, $n = 26$, spring participants, $n = 27$). They searched for answers to specific factual questions in educational psychology and earth science textbooks. Differences were found in search effectiveness (finding a correct answer to the question) only for the educational psychology text. Students tested in the fall (lower prior knowledge) performed significantly worse than students tested in the winter or spring, for questions related to content covered in the fall semester of the course. Likewise, students tested in the spring (higher prior knowledge) performed significantly better than students tested in the fall or winter, for content covered in the spring semester. Goldman and Duran (1988) found a similar effect of prior knowledge for undergraduate students searching in Educational Psychology and Earth Science textbooks; Rouet (2003) found the same results for undergraduates searching in geology and psychology hypertexts.

Byrnes and Guthrie (1992) compared search efficiency in human anatomy texts for 32 undergraduates who were low or high in prior knowledge about 4 organs. Students searched a 41-page researcher-designed 4-chapter booklet for answers to two questions, one about the digestive system (more familiar) and one about the excretory system (less familiar). One-half of the students received a booklet with the text organized in an unusual way (non-standard text; organized according to major concepts, e.g., “Transport of Materials”). Dependent measures were time to find the answer, number of chapters searched, and number of times students consulted the table of contents. Knowledge

differences emerged only with the standard text—high-knowledge students spent significantly less time finding the answer and searched significantly fewer chapters, but consulted the table of contents significantly less. All students in the non-standard text condition were inefficient searchers.

However, there is some contradictory evidence for the role of background knowledge from other studies of search behavior. Dreher and Brown (1993) and Dreher and Guthrie (1990), found no main effect for prior knowledge on searching. They did, however, find significant effects for efficiency of search on quality of answers to search questions.

Experimental intervention studies. Miyake and Norman (1979) compared question-asking for 60 undergraduates who they either trained about computer text editing commands (high PK) or did not train (low PK). Students thought out loud while reading an easy or hard version of a short manual explaining the text editor. Transcripts were coded for number of questions (adjusted for verbalizations), as well as interpretations and re-reading. There was a significant interaction between PK and text difficulty: low PK participants asked more questions about the easy text, while high PK students asked more questions about the hard text. Miyake and Norman conclude that “To ask a question, one must know enough to know what is not known” (p. 357).

Several reading comprehension strategy instruction intervention studies have compared the effects of an intervention on low- and high prior knowledge participants and failed to find effects of differences in background knowledge. That is, strategy *instruction* seems to be equally effective for low- and high-knowledge students. Schmidt et al. (1989) did not find differences in the effect of teaching a prior knowledge activation strategy on recall of explanations or descriptions from a science text between 9th grade

(“novice”) and 10th grade (“expert”) participants. Taylor and Beach (1984) similarly found no effects for prior knowledge for 114 above average 7th grade students when they studied the effectiveness of a summarizing strategy. Dole et al. (1991) likewise found no main effect or interactions for prior knowledge in a study of pre-teaching vs. prior knowledge activation strategy instruction with 5th grade students.

A further piece of evidence regarding the effect of background knowledge in use of strategies comes from a synthetic review of findings from learning strategy interventions (Weinstein, Husman, & Dierking, 2000). Weinstein et al. conclude that learning strategy interventions that are embedded in a subject-matter classes and reading materials have led to more transfer than those taught in separate stand-alone sessions using texts that are not connected to subject-matter classes. This suggests that there is a link between prior knowledge and strategy use.

Regression studies. Peverly et al., (2003) regressed text-specific background knowledge, study time, and student confidence ratings on scores on a written summarization measure with undergraduates. Background knowledge had a significant beta weight of .37 for students who were allowed to take notes while learning ($n = 41$), with an overall R^2 of .55.

Bråten and Samuelstuen (2004) found a significant effect of background knowledge on self-reported strategy use. They adapted the LASSI for a sub-sample of 269 Norwegian 10th grade students from the PISA reading study, measuring memorization, elaboration, organization, and monitoring. Background knowledge had a significant beta in all four regressions when entered together with reading goal conditions and interaction terms.

Correlational studies. Significant correlations between background knowledge and strategy use have occasionally been reported in the literature and are summarized in Table 4.

Table 4

Reported Correlations between Background Knowledge and Strategy Use, By Author

Study	Correlation	Participants	Text
Alexander, Murphy, Woods, Duhon, & Parker (1997)	.32 at pretest .15 at posttest	329 pre-service teachers	Educational psychology text
Artelt et al. (2001)	.24	622 15-year-old German students.	origin of the Earth text
Britton et al. (1998)	NSD	211 Air Force recruits age 17-25.	Vietnam War text
Kozminsky & Kozminsky (2001)	.66 across groups .47 academic students .44 semi-academic students .38 vocational students .34 LD students	205 9 th grade 33 academic 121 semi-academic 21 vocational 22 LD	Two existing Hebrew multiple-choice reading comprehension measures.

In summary, there is evidence from 6 experimental studies that background knowledge has a direct effect on use of reading comprehension strategies. However, five experimental studies failed to find evidence for such an effect. Five correlational studies were also found that were consistent with this effect. Due to the conflicting evidence for this path, it was selected as one of the paths to be tested in the variations of the IM model.

Path 3: Effect of background knowledge on inference. Background knowledge is expected to have an effect on inference because many inferences integrate

prior knowledge with information from the text. This line of research continues a topic that has been of interest since the early years of the “cognitive revolution” (see, e.g., Chi, Hutchinson, & Robin, 1989 on children’s knowledge about dinosaurs).

Experimental non-intervention studies. Fincher-Kiefer (1992) conducted 2 experiments, each of which compared 30 undergraduates who were low ($n = 10$), intermediate ($n = 10$) or high in baseball knowledge on a sentence completion task requiring inferencing. Students read a 714-sentence long description of a baseball game which contained sentences that were missing words (cloze-type sentences). The missing words completed global (text-wide) or local (sentence-level) inferences. In Experiment 1, students with higher knowledge made significantly more correct inferences and also had longer reading latencies, indicating that they were engaging in the time-consuming activity of inference making. Experiment 2 found the same results for correct inferences, but did not measure reading latency.

Franks (1997) compared the effect of prior knowledge on inference form (e.g., *modus tollens*) in three grades cross two studies. Experiment 1 included 21 fourth-grade and 30 seventh-grade students, and 26 college undergraduates, all of average reading ability. Participants read one-paragraph stories and answered questions requiring inferences, using a yes/no/can’t tell format. Background knowledge was manipulated by writing the stories about familiar topics (e.g., dogs and cats) or unfamiliar topics (e.g., a spaceship). Background knowledge had a significantly effect on more difficult types of inferences, but not on easier types. Experiment 2 included 40 7th-grade students and 40 college undergraduates, at a range of reading skill. Reading skills were tested with the California Achievement Test or the Nelson-Denny Reading Battery. Unfamiliar prior knowledge affected low-skill readers more at both ages. There were also complicated

interactions among prior knowledge, inference form, and age. For seventh graders, background knowledge had a significant effect on performance on all types of inferences. For college students, background knowledge had a significant effect on performance on more difficult types of inferences, but not on easier types.

Franks (1998) replicated the study with three older adult groups: college-aged ($n = 44$), middle-aged ($M = 38.5$ yrs, $n = 20$), and elderly ($M = 68.3$ yrs, $n = 20$), all with some college education. Background knowledge had a significant effect on performance on more difficult types of inferences, but not on easier types.

McNamara et al. (1996) measured the reading comprehension of students with different levels of background knowledge while reading text that was either more- or less-coherent and did or did not contain text signals referring to the text's macrostructure. Fifty-six students entering grades 7-10 completed prior knowledge measures including listing and describing parts of the heart, matching parts of the heart to a diagram, and answering fill-in-the-blank and short-answer questions. Students also sorted cards with biology terms into categories. Posttest measures included a 40-item short answer posttest and the sorting task. They found that high-knowledge students answered significantly more inference questions correctly than did lower-knowledge readers. High knowledge readers also answered significantly more inference questions correctly on low-coherence texts where they had to actively work to make sense out of the text; the opposite was true of low-knowledge readers.

Carr and Thompson (1996) measured prior knowledge for 48 fifth-grade non-learning disabled (NLD) and 8th grade NLD and learning disabled (LD) students as part of a strategy intervention study. All students answered more inferential questions correctly if they had more background knowledge.

Schneider et al. (1989), by contrast, found no effect of background knowledge on inference in Experiment 1 with 3rd, 5th and 7th grade German students.

Experimental intervention studies. Blanc and Tapiero (2001) presented 72 college undergraduates with a high-specific preview ($n = 24$), a low-specific preview ($n = 24$), or an irrelevant preview ($n = 24$) before they read a text requiring spatial inferences. After reading each sentence, participants were presented with test sentences to determine whether they had drawn the relevant inferences. Participants with more background knowledge drew more accurate inferences and also had shorter reading times.

Vidal-Abarca, Martinez, and Gilalbert (2000) gave 64 8th-grade students one of four versions of a text about Russian history: an unmodified version, a version re-written to require fewer inferences (argument overlap; AO), a version re-written with more background knowledge that was designed to increase students' inferences (causal connectionist; CC), and a version that combined these two (AO+CC). On a researcher-developed 7-question open-ended inference task, both background knowledge versions (CC and AO+CC) resulted in significantly higher inference scores.

Roberts (1988) found that pre-teaching of soccer concepts to 9th grade college preparatory students (via self-study) yielded significantly higher inference scores than did a prior knowledge activation strategy.

Stahl et al. (1989, see p. 76), however, failed to find a significant effect for prior knowledge on inference. The researchers pre-taught information about an Amazonian tribe to 6th grade students. Students who received pre-teaching that was relevant to the main ideas of the text performed no better on Cloze (Experiment 2) and sequencing and

importance rating (Experiment 3) measures than did students who received irrelevant pre-teaching.

Regression studies. Peverly et al. (2003) regressed text-specific background knowledge, study time, and student confidence ratings on inference scores from a free recall and multiple-choice measure for undergraduates. Background knowledge had a significant beta weight of .34 and an $R^2 = .55$ for students who were allowed to take notes while learning ($n = 41$).

Callahan and Drum (1984; see p. 37) found that prior knowledge accounted for a significant proportion of variance in inference and in cloze tasks ($R^2 = .45$ and $.50$, respectively) for 5th and 6th grade students; prior knowledge was the only predictor to enter the regression equations.

Correlational studies. Graesser and Bertus (1998) found a significant correlation between background knowledge and inference of $r = .58$ for younger adults ($M = 22$) and $.43$ for older adults ($M = 67$). The measured background knowledge with a composite of the WAIS Information and Vocabulary subtests and inference with a composite of Raven's Progressive Matrices and the ETS Inference Test. See Graesser et al. (1994) for a review of background knowledge and inference.

In summary, there is evidence from 8 experimental studies that background knowledge has a direct effect on inference. Five regression and correlational studies were also found that are consistent with this effect. Studies were conducted with students in 5th grade through older adults. Most measures were researcher-designed multiple-choice tests, but included sentence recognition, sentence reading times, and a nationally-normed test.

Path 4: Effect of strategies on comprehension. The effect of strategies such as summarizing, self-questioning, using graphic organizers, and search on comprehension has been a major research program since the 1980s. Think-aloud studies in the late 1970s and early 1980s showed differences between high- and low-comprehending students in their ability to enact strategies. Researchers then designed and tested the effects of reading comprehension strategy instruction interventions.

Experimental non-intervention studies. Several experimental studies provide evidence that students with greater knowledge of strategies perform better at reading comprehension tasks. Lau and Chan (2003) found significant differences among Chinese 7th grade students between 83 good comprehenders and 76 poor comprehenders on a researcher-developed strategy use measure. Good comprehenders were significantly better on all measures, which tested deleting irrelevant sentences, identifying the main idea of a paragraph, writing topic sentences, and identifying errors in short paragraphs.

Schneider et al. (1989) in Experiment 1 found that 3rd, 5th, and 7th grade students who were better able to detect contradictions (a metacognitive monitoring strategy) were better able to recall text details. However, in Experiment 2 they found no differences between good and poor readers in scores on a questionnaire about knowledge of comprehension strategies. Körkel (1987, cited in Schneider, 1993) found that text recall was higher for students with high scores on a questionnaire measuring declarative metacognitive knowledge (knowledge about strategies).

Meyer et al. (1980) compared 102 good, average and poor 9th grade readers (as measured by Stanford reading tests) on problem/solution and compare/contrast texts. Better readers used the strategy of text structure significantly more during immediate and delayed free recall than did poor readers.

Reynolds, Shepard, Lapan, Kreek and Goetz (1990) found that 25 10th-grade better comprehenders (as measured by the Nelson-Denny comprehension test) were significantly better able to use the strategies of monitoring and identifying the main ideas in a 36-page science text than were 20 less-proficient comprehenders.

Experimental intervention studies. Perhaps more than any other form of evidence, reading comprehension strategy intervention studies have demonstrated the effect of strategy use on reading comprehension (for reviews, see NRP, 2000; Pressley, 2000; Rosenshine & Meister, 1997). In the interest of brevity, the results of 46 reading comprehension strategy interventions with middle school and high school students are summarized in Appendix A.

For the strategy of graphic organizers, 4 middle school and 7 high school studies were identified that showed significant effects on comprehension for one or more outcome measures (see Appendix A for details of all of the studies). For the strategy of prior knowledge activation, 4 middle school and 7 high school studies were identified (see Table 2). For the strategy of question generation, 5 middle school and 4 high school studies were identified. For the strategy of story structure, 3 middle school and 1 high school study were identified. For the strategy of summarizing, 6 middle school and 2 high school studies were identified. For multiple strategy instruction (e.g., Reciprocal Teaching), 6 middle school and 3 high school studies were identified. In all, 55 studies in Table 2 and Appendix A with middle and high school students show significant effects of strategy instruction on at least one reading comprehension measure.

Path analysis studies. Artelt et al. (2001, see p. 51) used a researcher-developed measure of knowledge of metacognitive strategies, a researcher-developed multiple-choice questions measuring prior knowledge about the origin of the

Earth, and a measure of learning from text (unlike in most reading comprehension studies, the text was not available to participants when answering questions about it). They found the strategy knowledge measure loaded a significant .42 on a multiple-choice comprehension test using an SEM model. Fit of the model was excellent; RMSEA = .003, CFI = 1.0.

Alexander et al. (1997) tested an SEM model for knowledge, interest, and strategies on recall from educational psychology text with 329 undergraduates. They found a significant loading of strategies on recall of .28 at pretest and .33 at posttest. The model fit the data moderately well, with an AGFI of .90.

Schoonen, Hulstijn and Bossers (1998) used an SEM model to test the contribution of strategies and vocabulary to reading comprehension for Dutch students ($n = 132$ 6th grade, $n = 178$ 8th grade, and $n = 178$ 10th grade). They used a 4-part researcher-developed metacognitive and cognitive strategy self-report measure, including knowledge of oneself as a reader, of text characteristics, of specific reading strategies, and of goals and criteria for comprehension. In 6th grade, none of the four aspects of strategy knowledge made a significant contribution to comprehension. In 8th grade, text characteristics, strategies, and goals had significant loadings (.26, .19, and .21 respectively), and in 10th grade text characteristics had a significant loading of .42. Their model explained 62-65% of the variance in scores on a standardized reading comprehension measure, and had an excellent fit to the data (chi-square tests for all three grade levels were non-significant, GFI = .99-1.00).

Regression studies. Cain et al. (2004; see p. 61) found that scores on a researcher-developed error detection test of metacognitive monitoring explained a

significant 3% of variance in reading comprehension for 4th and 6th grade students, after accounting for word reading accuracy, vocabulary, IQ, and working memory.

In a multi-component study, Ehrlich, Kurtz-Costes, & Loridant (1993) studied 127 French seventh-grade students, 64 good readers (70th percentile) and 63 poor readers (30th percentile) as measured by a standardized test. They regressed strategy knowledge (measured with a questionnaire about metacognitive knowledge about reading), motivation (academic self-concept and attributions), and single word recognition on comprehension. Good readers had significantly higher scores on all measures except attributions. For good readers, only academic self-concept explained a significant proportion of variance in reading comprehension, and the total $R^2 = .15$. For poor readers, only word reading explained a significant proportion of variance in reading comprehension, and the total $R^2 = .09$. The small sample size relative to the number of variables, and the predictors chosen (e.g., vocabulary was not included) gives this study limited power to detect significance.

Saarnio, Oka, and Paris (1990) conducted a multi-component study with 426 students, 213 from 3rd grade and 216 from 5th grade. They measured metacognitive strategy knowledge (measured by a questionnaire), inference (measured with a cloze task), memory for just-read text, self-perceptions regarding reading, and decoding. They entered those variables with significant correlations with reading comprehension (as measured by the Gates-MacGinitie comprehension test) into a regression equation. When entered last, strategy knowledge explained a significant but small 4% of the variance in 5th grade reading comprehension. Overall, the regression explained 46% of the variance in 5th grade comprehension.

Think-aloud studies. In a think-aloud study with 8 9th-grade students at a range of teacher-rated reading levels reading short stories by William Faulkner, Rogers (1991) identified a number of specific strategies used by high school readers. These included summarizing, elaboration, monitoring (both expressions of understanding and lack of understanding), hypothesizing, and evaluating.

Wade, Trathen, and Schraw (1990), however, found no significant differences in comprehension across six different strategy use clusters (e.g., “the good strategy user,” “the memorizer”). They asked 67 undergraduates from education classes and a strategy skills class to verbalize strategies while studying, and scored students on the use of 14 strategies (e.g., paraphrasing in notes, rereading).

Four additional middle school and eight additional high school think-aloud studies showing evidence of strategy use were identified and are summarized in Appendix B. These studies used a variety of types of texts and only sometimes compared high- and low-comprehending or older and younger readers. Verbalizations during think-alouds included evaluating what was read, hypothesizing/predicting, monitoring, rereading, relating what was read to prior knowledge, summarizing/paraphrasing, using text structure, and visualizing. Because of previously-discussed data analysis problems with frequency data, no patterns can be discerned for high school students’ use of strategies. For comprehensive reviews of strategy use in think-aloud studies see Pressley and Afflerbach (1995) and Afflerbach (2000).

Correlational studies. In addition to strategy instruction intervention, multi-component, and think-aloud studies, there is both weak and strong correlational evidence for the relationship between strategies and reading comprehension, which is summarized in Table 5.

Table 5

Reported Correlations between Strategies and Comprehension, By Author

Study	Correlation	Participants	Strategies
Artelt et al (2001; see p. 51)	.51	622 15-year-old German students	Metacognitive strategies
Bråten and Samuelstuen (2004)	.06-.23	269 10 th -grade students randomly sampled from the Norwegian PISA sample, poor readers were excluded	Memorization, elaboration, organization, and monitoring
Cain et al. (2004)	.48 4 th grade .54 6 th grade	92 4 th grade and 80 6 th grade students from a mixed SES school followed longitudinally, extremely good and extremely poor readers were excluded	Metacognitive monitoring (error detection)
Chan (1994)	.22 (ns)	101 ninth-grade (25 with learning difficulties), 133 seventh-grade, and 104 fifth-grade students	25 Likert-type strategy knowledge and use questions (e.g., “Mary knows that a paragraph often has a key sentence . . . How helpful is it? . . . How often do you read this way?, p. 339)
Kozminsky & Kozminsky (2001)	.56 - .65 (varied across groups: .77 for academic students, .44 for semi-academic, .49 for vocational, and .36 for LD)	205 9 th grade 33 academic 121 semi-academic 21 vocational 22 LD	Multiple-choice strategy use questions (for summarizing, clarifying, questioning, and predicting, e.g., “One of the following sentences is a good summary.” p. 204)

Study	Correlation	Participants	Strategies
McBride-Chang, Manis, Seidenberg, & Custodio (1993)	.40	49 non-reading disabled 5 th -9 th grade students.	Metacognitive monitoring
van Gelderen et al. (2004)	.85	281 Dutch 8 th graders across a range of academic tracks	80-item questionnaire with correct/incorrect items (e.g., "It is sensible to put most effort into memorizing," p. 22).

van den Broek, Lynch, Naslund, Ievers-Landis, and Verduin (2003) found significant increases in the ability to identify main ideas associated with increases in comprehension across 757 students in 3rd, 6th, 9th and 11th grades reading narrative text.

In summary, there is evidence from 50 experimental studies that knowledge of and use of reading comprehension strategies has a direct effect on reading comprehension. Twenty-seven path analysis, regression, think-aloud, and correlational studies were also found consistent with this effect. These findings apply across self-report and observational studies, reports of both knowledge of and use of strategies, and across metacognitive and cognitive strategy use.

Path 5: Effect of strategies on inference. In reading comprehension strategy instruction interventions, outcomes have sometimes been reported separately for inferential and literal questions. It is therefore possible to separate out the effect of strategies on inference. It may be that strategic processing is only needed for difficult comprehension problems such as inferential questions. If that is the case, then strategies should have most of their effect on comprehension indirectly, via inference, and not directly on comprehension.

Experimental non-intervention studies. Does better knowledge of reading comprehension strategies lead to more accurate inferencing in reading? Schneider et al. (1989) in Experiment 1 found that 3rd, 5th, and 7th grade students who were better able to use the metacognitive strategy of detecting inconsistencies were also significantly better able to make inferences.

Experimental intervention studies. Several strategy instruction intervention studies have measured outcomes of the intervention with inferential questions, as well as literal comprehension questions. Baumann (1984) taught 61 6th grade students at a range of reading skills to summarize social studies texts. Treatment students answered significantly more multiple-choice comprehension questions that required inferences than did students in either a basal plus question answering control group or a no-treatment control group. Carnine, Kameenui and Woolfson (1982) found similar results for a summarizing treatment with 36 5th grade students.

Hansen and Pearson (1983) taught good and poor 4th grade readers to ask themselves questions while they read. They found that treatment students scored higher on comprehension questions requiring inferences than did control students.

Carr and Thompson (1996) trained 48 5th grade non-learning disabled (NLD) and 8th grade NLD and learning disabled (LD) students to summarize stories with familiar or unfamiliar content. All students improved performance on inference questions as a result of the treatment, but LD students benefited more than did NLD.

Mathewson (1989) compared prediction, note-taking, and summarizing treatments for 11th grade average and high-achieving students ($n = 24$) in social studies. For amount written, the prediction and note-taking treatments were significantly more effective than summarizing. For causal links (an inference measure) and relevance, the prediction

treatment was significantly more effective than the note-taking or summarizing treatments.

By contrast, Graves et al. (1983; see p. 40) found no effect of advance organizer strategy instruction on inference for 8th grade struggling readers, although they did find an effect on literal comprehension.

Path analysis studies. Britton et al. (1998; see p. 41), in their SEM study of Air Force recruits reading a Vietnam War text, found that metacognitive ability loaded a significant .37 on inferencing. The model had a slightly less-than-ideal fit: CFI = .94, RMSEA = .05.

Correlational studies. Cain et al. (2004; see p. 61) found significant correlations of .27 for 4th graders and .26 for 6th graders between scores on an error detection measure of metacognitive monitoring and answers to inference questions that followed short passages. Walczyk (1990) found a significant correlation of .53 for 4th graders between similar, researcher-developed measures.

In summary, there is evidence from 6 experimental studies that reading comprehension strategies have a direct effect on inference. Three path analysis and correlational studies were also found that were consistent with this effect. Studies were conducted with students from 3rd grade through young adults. Texts included narrative and expository (social studies) passages; measures were researcher-developed multiple-choice tests. As with the effect of reading comprehension strategies on comprehension, this finding holds for both metacognitive and cognitive strategy use.

Path 6: Effect of word reading on comprehension. A body of research indicates that word reading continues to have an effect on reading comprehension into the later grades, even though it might be thought to have “disappeared” for non-Reading

Disabled students. Word reading accuracy and/or fluency is expected to play a role in comprehension because it draws attentional resources away from effortful processes such as comprehension strategies (Samuels, 1994).

Experimental non-intervention studies. Haines and Leong (1983) compared high- and low-comprehending students in 4th, 6th, and 8th grades ($n = 72$) on accuracy and latency in reading regular and irregular real words, as well as pseudowords, in isolation. They found a significant main effect for comprehension level; high-comprehenders read all types of words more accurately and faster.

Leach, Scarborough and Rescorla (2003) found that some 4th and 5th grade students with reading comprehension problems had word reading difficulties, while some had a combination of word reading and comprehension strategy deficits, and others had “pure” comprehension deficits. Buly and Valencia (2002) found very similar results with 5th grade students who had failed the Washington State 4th grade literacy assessment. Catts, Hogan and Fey (2003) found similar results with 4th grade students who had shown word reading or listening comprehension problems in Kindergarten.

Curtis (1980) studied 40 5th-grade students and found that less-skilled comprehenders (as measured by the Diagnostic Reading Scales) read significantly fewer real words and pseudowords per minute than did more-skilled comprehenders.

Smiley, Pasquale, and Chandler (1976) tested 18 good and poor 7th grade readers’ (as measured by the California Test of Basic Skills) ability to read frequent and infrequent real words and non-words. All students had more difficulty with non-words than with infrequent real words, and more difficulty with infrequent real words than with frequent real words, and the differences across groups for real words were significant. Poor comprehenders read more words incorrectly.

Hood and Dubert (1983) gave 9th grade students both written and oral vocabulary tests and compared the results to the students' reading comprehension. Students completed one form of the Iowa Tests of Educational Development vocabulary subtest in its written version (ITED-V) and the alternative form in a tape-recorded version (Prompted-V), as well as the Iowa comprehension subtest. In a multiple regression with reading comprehension as the dependent variable, Prompted-V was entered first, and then ITED-V was entered. ITED-V accounted for an additional 13.4% of variance for 338 students from one suburban school, and an additional 5.7% of variance for 72 students from one rural school. Approximately 10% of the students showed a discrepancy between Prompted-V and ITED-V of more than twice the standard error for ITED-V; that is, they showed decoding difficulties.

Martino and Hoffman (2002) found that low-comprehending college students (as measured by their ACT scores) had lower Woodcock Word Identification scores than did high-comprehending students (scores of 97 and 100, respectively), but non-significantly lower Word Attack scores.

Experimental intervention studies. Word reading instructional programs have only sometimes been found to have an effect on reading comprehension with older children. The National Reading Panel reviewed 66 experimental studies on word reading interventions and found a significant average effect size of .51 for reading comprehension across 11 studies with children in K-1st grade; a significant average effect size of .32 across 9 studies with reading disabled children in 2nd-6th grades; but a non-significant average ES of .12 across 11 studies with children in 2nd-6th grades overall (NRP, 2000). One explanation may be that word reading interventions were too short to have an impact on comprehension. Another explanation may be that readers with word

reading problems also have vocabulary, background knowledge, inference, and other comprehension strategy problems that also interfere with comprehension.

Breznitz (1997) found that testing students on computer-based text at a faster rate than they normally read led to significantly more correct responses to comprehension questions for students in a longitudinal study through grade 5 ($N = 81$). The largest differences were seen when decoding was more error-filled; that is, speeding up slow decoders helped them to comprehend better.

Tan and Nicholson (1997) used flashcards to increase fluency for below-average 2nd-5th grade students from a whole language school. There were 18 students above 3rd grade in the study. Students trained in tutoring sessions with either words or phrases significantly outperformed students who were given vocabulary tutoring only. They found significant increases in students' comprehension of grade-appropriate texts, using verbal answers and recall of main ideas and supporting facts as outcome measures.

Bourassa, Levy, Dowin and Casey (1998) found that 4th grade students with poor reading accuracy ($n = 24$) could be trained to read faster using repeated readings, and that this significantly improved their reading comprehension compared to untrained controls.

On the contrary, Bryant, Vaughn, Linan-Thompson, Ugel, Hamff and Hougen (2000) provided an integrated decoding, fluency, and reading comprehension strategy instruction intervention to 29 average-achieving, 17 low-achieving, and 14 reading disabled 6th grade students. Only the students with reading disabilities showed significant improvement in word reading, and there were no significant gains in comprehension for any group.

Path analysis studies. In an SEM study, Artelt et al. (2001, see p. 51) found that decoding speed loaded a significant .26 on a multiple-choice

comprehension test (fit of the model was excellent; RMSEA = .003, CFI = 1.0).

However, with an outcome measure of learning from text (the text was not available to participants when answering questions about it), decoding speed loaded a non-significant .03 (the model had a slightly less-than-ideal fit: CFI = .94, RMSEA = .05).

Singer and Crouse (1981) found a significant *indirect* effect of decoding on comprehension via vocabulary ($\beta = .40 \times .62 = .25$) for 127 6th grade students in a path model that included vocabulary, letter discrimination, non-verbal IQ, and a cloze measure of inference. The direct effect of word reading accuracy on reading comprehension was non-significant.

Regression studies. Curtis (1980; see p. 65) regressed DRS comprehension on listening comprehension and pseudoword reading speed (in words per minute) for 40 5th-grade students. She found that pseudoword reading speed accounted for a small but significant 3% of the variance in comprehension.

Bell and Perfetti (1994, see p. 42) regressed word reading and other components on reading comprehension for college students, and found that word reading accounted for a significant 11% of variance when entered last. However, Haenggi and Perfetti (1994, see p. 42) did not find that word reading made a significant contribution for undergraduates.

Cunningham, Stanovich and Wilson (1990) conducted a multi-component study with 76 college undergraduates, measuring word reading accuracy, working memory, vocabulary, visual processing, listening comprehension, IQ, and print exposure on reading comprehension. They entered those variables with significant correlations with reading contribution into a regression equation. The components together accounted for 78% of the variance in comprehension on the Nelson-Denny test. They found that word

reading accounted for a significant 7.3% of variance in reading comprehension when entered last into the regression.

Jackson and McClelland (1979) measured word reading accuracy, working memory, vocabulary, visual processing, verbal aptitude, phonemic awareness and listening comprehension for 52 undergraduates. They entered those variables with significant correlations with reading contribution into a regression equation. Word reading contributed significantly only for a sub-sample of the 12 fastest and 12 average readers.

Levy and Hinchley (1990) measured word reading accuracy, working memory, vocabulary, rapid naming, phonemic awareness, memory for text, and IQ on Gates-MacGinitie reading comprehension for 345 students in 3rd-6th grades ($n = 88$ 3rd grade, $n = 92$ 4th grade, $n = 82$ 5th grade, and $n = 83$ 6th grade). They entered those variables with significant correlations with reading contribution into a regression equation. Word reading made a significant contribution to variance in reading comprehension in all four grades.

Stage and Jacobsen (2001) measured oral reading fluency in September, January, and May of 4th grade. Using HLM, they found that both average fluency (intercept) and slope (how fast students' fluency grew) significantly predicted scores on the Washington Assessment of Student Learning reading test given at the end of 4th grade.

Baddeley, Logie, Nimmo-Smith and Brereton (1985) regressed word reading accuracy, vocabulary, and working memory on reading comprehension in two experiments. In Experiment 1, with 51 adults ages 18-66 ($M = 40$), $R^2 = .44$ and in Experiment 2, with 107 adults ages 19-60 ($M = 41$), $R^2 = .43$, word reading was a significant predictor of reading comprehension.

Ehrlich et al. (1993, see p. 59) found that word reading accuracy explained a significant proportion of variance in comprehension for poor 7th grade readers. However, the total R^2 for poor readers was only .09.

Shankweiler et al. (1996) investigated relationships among nonsense word reading (using a measure from Olson, Fosberg, Wise and Rack), comprehension (using the Fast Reading subtest of the Stanford Diagnostic Reading Test), and print exposure (using Stanovich's Magazine Recognition Test). Participants were 86 9th grade students from a public high school from a range of achievement groups. The correlation between nonsense word reading and comprehension was a significant .45. Even after accounting for print exposure, decoding added an additional significant 10.7% to the variance in comprehension accounted for.

McBride-Chang et al. (1993, see p. 62) measured print exposure, IQ, word and non-word reading, memory for just-read text, metacognitive strategies, and spelling for 49 non-reading disabled students in grades 5-9. A regression of age, word identification, and print exposure accounted for a significant $R^2 = .19$ in reading comprehension, but no single predictor was significant.

Saarnio et al. (1990, see p. 59) found that word reading explained a significant 5% of variance in reading comprehension for 3rd and 5th graders when entered last.

Think-aloud studies. One think-aloud protocol study in reading identified word-level strategies used by high school students. Fehrenbach (1991, see p. 44) identified "word pronouncing concern" as a word-related strategy. Word pronouncing concern was used more often by poor readers than by good readers.

Correlational studies. In a meta-analysis of 17 studies, Gough, Hoover and Peterson (1996) found mean correlations of .61 between word reading

accuracy and comprehension for children in 1st-2nd grades, .53 for children in 3rd-4th grades, .48 for 5th-6th grade, and .39 for college students, with a significant developmental trend for decreasing correlations over time. A number of studies in addition to those reported by Gough et al. have found significant correlations between word reading accuracy and/or speed and comprehension. These are summarized in Table 6.

Table 6

Reported Correlations between Word Reading and Comprehension, By Age (Studies Not Included in Gough et al. [1996] Meta-Analysis)

Study	Correlation	Participants	Word reading measure
Cain et al.(2004; see p. 61)	.40 4 th grade .19 (ns) 6 th grade	92 4 th grade and 80 6 th grade students from a mixed SES school followed longitudinally, extremely good and extremely poor readers were excluded.	Neale word reading accuracy
De Soto & De Soto (1983)	.60 for pseudoword accuracy -.53 for pseudoword time -.48 for real word time	134 4 th grade students	Real word and pseudoword accuracy and speed
Jenkins, Fuchs, van den Broek, Espin and Deno (2003)	.50 for accuracy .83 for speed (words per minute)	113 4 th grade students	Speed and accuracy
Nagy, Berninger, Abbott, Vaughan and Vermuelen (2003)	.50 for Word Attack .61 for Letter-Word Identification	4 th -grade students who had failed a writing exam	Word Attack and Letter-Word Identification

Study	Correlation	Participants	Word reading measure
McBride-Chang et al. (1993, see p. 62)	.41	49 5 th -9 th grade students	Word identification
van Gelderen et al. (2004, see p. 62)	-.22	281 Dutch 8 th graders across a range of academic tracks	Latency for judgments of whether a letter string was a real word or not
MacDonald & Cornwall (1995)	ns	24 students who had participated in a kindergarten PA study [T1] at age 17 [T2]	Word Attack
Saiegh-Haddad (2003)	ns	50 undergraduate students	Oral reading fluency in L1
Dixon, LeFevre, & Twilley (1988)	.28	95 undergraduate students	Speed
Stanovich & West (1989)	.51	61 undergraduate students	Word identification

In summary, there is evidence from 9 experimental studies that word reading has a direct effect on reading comprehension. In addition to the studies in the Gough et al. (1996) meta-analysis, 22 path analysis, regression, think-aloud, and correlational studies were also found that were consistent with this effect. This finding holds true across fluency and accuracy measures, and across real word and pseudoword reading.

Path 7: Effect of vocabulary on inference. Although rarely researched, vocabulary might be expected to have an effect on inference for the same reason that vocabulary is expected to have an effect on strategies: without understanding the meaning of a word, the reader cannot draw logical conclusions using what was read.

Experimental non-intervention studies. Stahl et al. (1989; see p. 76) in Experiment 2 with 92 6th-grade students, found a significant effect of vocabulary difficulty on cloze performance for function words (articles, pronouns, conjunctions, and modal verbs). In Experiment 3, involving 99 6th-grade students, showed a significant main effect of vocabulary difficulty on identifying causal relations.

Experimental intervention studies. Medo and Ryder (1993) pre-taught relevant vocabulary to 8th grade students. Participants were matched on prior knowledge about content from the reading passage—volcanoes. Treatment ($n = 31$) and control ($n = 31$) students read a 429-word passage about volcanoes from a trade book and answered 11 questions that required students to make causal connections (inferences and elaborations). Treatment students answered significantly more causal connection questions correctly than did control students.

Kameenui, Carnine, and Freschi (1982, see p. 78) found that 4th-6th grade students reading social studies text with easy-vocabulary passages scored significantly better on inference questions than those who read passages with difficult vocabulary. In addition, students who received vocabulary instruction also scored significantly better on inference questions than those who received no instruction.

Carney, Anderson, Blackburn and Blessing (1984) found that 5th-grade students who were pre-taught vocabulary subsequently performed significantly better on inference questions than did students who were not pre-taught.

Path analysis studies. Singer and Crouse (1981; see p. 68) found that vocabulary loaded a significant .39 on a cloze measure of inference for 127 6th grade students in a path model that included decoding, letter discrimination, non-verbal IQ, and comprehension.

Regression studies. Dixon et al. (1988, see p. 72) regressed vocabulary and working memory on inference for college students, and found that vocabulary accounted for a significant 27% of variance when entered last, but their overall R^2 was only .17.

Correlational studies. Cain et al. (2004; see p. 61) found significant correlations of .32 for 4th graders and .48 for 6th graders between scores on the Gates-MacGinitie vocabulary subtest and a researcher-developed multiple-choice test of inference questions that followed a short passage.

Davey (1987) found a significant correlation of .55 between vocabulary knowledge and inferential comprehension for 5th-6th grade good readers, and a significant correlation of .47 between vocabulary knowledge and inferential comprehension for 10th grade poor readers.

Walczyk and Taylor (1996) tested Walczyk's Compensatory-Encoding Model (C-EM); that readers with poor lexical access and/or working memory deficiencies can compensate for them by using various control strategies such as slowing down, looking back in text and re-reading, or pausing to integrate propositions. Walczyk and Taylor found these differences with 109 undergraduate students who read 6 short texts on a computer using a moving-window paradigm. Participants who were slower at naming single words also looked back more in text. They found significant correlations of .44 between vocabulary and anaphor reference latency, and -.24 between vocabulary and anaphor reference accuracy, indicating that high vocabulary knowledge is associated with quick response times. Walczyk (1990) found a significant correlation of .35 between the Iowa Basic Skills Test vocabulary subtest and a researcher-developed inference measure for 37 4th grade students.

In summary, there is evidence from 4 experimental studies that vocabulary knowledge has a direct effect on students' ability to draw inferences in reading. Six regression, path analysis, and correlational studies were also found consistent with this effect. Inference measures included researcher-constructed multiple choice, cloze, and anaphor resolution.

Path 8: Effect of vocabulary on comprehension. Vocabulary is expected to make a large contribution to reading comprehension because readers cannot make sense of text if they cannot access the meanings of the words, and thereby activate related background knowledge; Stahl (1998) refers to this as the instrumentalist hypothesis.

Experimental non-intervention studies. Graves et al. (1980) investigated the effect of vocabulary on reading comprehension for 7th through 12th grade students. Teachers divided students into low and high reading ability groups, and students took a researcher-designed multiple-choice definition vocabulary test. There was a significant main effect for reading ability: high-ability students answered approximately 8% more vocabulary questions correctly than did low-ability students.

A second experimental approach to investigating the effects of vocabulary knowledge on comprehension is comparing comprehension on texts written with easy and difficult vocabulary. Freebody and Anderson (1983a) found that 6th grade students who read social studies passages written with high-frequency vocabulary showed better comprehension than those who read passages where difficult vocabulary had been substituted. Experiment 1 ($N = 84$) showed significant main effects of vocabulary on oral recall and written summarization measures of comprehension. Eight-eight students participated in Experiment 2, which showed a significant main effect of vocabulary on sentence verification. Freebody and Anderson (1983b) found similar results for 79 sixth-

grade students reading social studies text in Experiment 1, which showed a main effect for vocabulary on recognition and in Experiment 2 with 71 6th grade students which showed a main effect on recall.

Stahl et al. (1989) found similar results with other samples of 6th grade students, also reading social studies text. Measures of comprehension across three experiments included literal and inferential multiple-choice, sentence verification recall, cloze, ordering the events in the passage, multiple-choice causal questions, and rating the importance of sentences. In Experiment 1, with 90 6th grade students, there was a significant main effect on passage recall for vocabulary difficulty. Experiment 2, with 92 6th grade students, showed the same significant main effect on cloze performance for vocabulary. Experiment 3, involving 99 6th grade students, showed a significant main effect for vocabulary on event ordering.

With regard to inferring the meaning of unknown words, Lau and Chan (2003) found significant, almost two-fold differences between good and poor 7th-grade comprehenders in scores on a researcher-developed measure.

In the area of morphology, Nagy et al. (1993) considered the relationship between knowledge of derived words (e.g., roarer) and the words from which they are formed (stem words; e.g., roar). Participants were 254 students in 4th, 7th, and 9th-12th grades. Students took a researcher-developed test of prefix and suffix knowledge, in which they had to choose the correct usage for a nonsense word such as “powderize” (derivative items). In this example, answer choices used the word as verb (correct; “First they had to find a way to powderize the rock” p. 161); adjective (incorrect), or noun (incorrect). Students also answered parallel items with stem words (e.g., “First they had to find a way to smash the rock” p. 161). Thus, this was a test of syntax, or word form, knowledge.

Students in 7th grade and above answered significantly more stem items and derivative items than did 4th grade students, even after adjusting for reading comprehension. Nagy et al. identified a group of students who performed well on stem items but not on derivative items, suggesting that morphological knowledge makes a unique contribution to vocabulary scores.

Mahony (1994) modified and added to Nagy's measures to create the Morpheme Sensitivity Test, which measures knowledge of syntax, relationships between word pairs (e.g., angel, angelic), and silent letter pronunciation (e.g., sign, signature) for suffixes. Three different experiments were conducted, with 26 college undergraduates, 24 ninth-grade students, 24 advanced placement 11th grade students, and 56 youth literature (low-reading) students in 10th-12th grades. On all measures, youth literature students scored the lowest, followed by 9th grade, AP, and undergraduates. Students with better reading comprehension had significantly better knowledge of suffixes in all three areas: syntax, relationships between word pairs, and silent letter pronunciation.

Experimental intervention studies. Intervention studies have been somewhat successful at increasing students' vocabulary knowledge and reading comprehension (NRP, 2000). In general, interventions that allow students to use new words multiple times in different contexts have been most successful; those that teach only conceptual relationships are somewhat successful, and those that teach students dictionary-type definitions have not been very successful (Nagy & Scott, 2000).

McKeown, Beck, Ohmanson and Perfetti (1983) taught 104 difficult new vocabulary words to 4th grade students in 75 thirty-minute lessons. Students used the words in multiple contexts (e.g., reading, writing, games) repeatedly over 5 months. Treatment students scored significantly higher than controls on a vocabulary test;

category decision task; comprehension test; and vocabulary, length, and structure of story recall.

Kameenui et al. (1982) compared three vocabulary-manipulation conditions and two vocabulary instruction conditions using researcher-constructed passages. Experiment 1 involved 60 4th ($n = 16$), 5th ($n = 21$) and 6th ($n = 23$) grade students. Students who read easy-vocabulary passages scored significantly better on inference questions and recalled significantly more segments involving difficult vocabulary than those who read passages with difficult vocabulary. Students who received vocabulary instruction also scored significantly better on the same measures than those who received no instruction. In Experiment 2, 60 5th and 6th grade students participated in the same treatments, and showed the same results.

Carney et al. (1984) compared 5th grade students who were pre-taught vocabulary that was important to a social studies textbook passage with control students who were not pre-taught ($N = 25$). They found a significant effect of pre-teaching on a researcher-constructed comprehension posttest including both literal and inferential questions.

Wixson (1986) found similar results for 120 above-average 5th grade students reading two historical fiction passages. Students who were pre-taught vocabulary that was central to the stories had significantly better comprehension than those who were pre-taught unimportant vocabulary.

Margosein, Pascarella and Pflaum (1982) compared two methods of teaching vocabulary to 7th and 8th grade students, and its impact on comprehension. Twenty-one students received semantic mapping vocabulary instruction and 23 received a context clues treatment. Semantic mapping participants had significantly higher posttest scores on the Gates-MacGinitie vocabulary test and a matching test than did the context clues

group. An ANCOVA showed that mapping students performed significantly better on the Gates-MacGinitie comprehension posttest than did context students.

Path analysis studies. Schoonen, Hulstijn and Bossers (1998) used an SEM model to test the contribution of vocabulary and reading comprehension strategies to reading comprehension for Dutch students ($n = 132$ 6th grade, $n = 178$ 8th grade, and $n = 178$ 10th grade). Vocabulary had a significant loading of .65 in 6th grade, .32 in 8th grade, and .58 in 10th grade. Their model explained 62-65% of the variance in scores on a standardized reading comprehension measure, and had an excellent fit to the data (chi-square tests for all three grade levels were non-significant, GFI = .99-1.00).

Singer and Crouse (1981, see p. 68) found that vocabulary made the largest contribution to comprehension (total effect = .71) for 127 6th grade students in a path model that included vocabulary, a cloze measure of inference, letter discrimination, non-verbal IQ, and decoding.

Regression studies. Several multi-component studies have also identified vocabulary as playing a role in reading comprehension. Baddeley et al. (1985, see p. 69) in Experiment 1 regressed vocabulary and other components on reading comprehension for adults ages 18-66 and found that it accounted for a significant 7.1% of variance when entered last.

Levy and Hinchley (1990) found, likewise, that vocabulary accounted for a significant proportion of variance in 3rd-6th grade students' reading comprehension. Bell and Perfetti (1994, see p. 42), Cunningham et al. (1990, see p. 68), and Dixon et al. (1988, see p. 72) found similar results with college students. However, Jackson and McClelland (1979, see p. 69) failed to find a significant contribution of vocabulary to reading comprehension with a similar college sample.

Stahl et al. (1991) measured baseball-related vocabulary and comprehension of a baseball passage for 159 10th grade students (testing at the 64th percentile on the Nelson-Denny). Based on prior research, Stahl et al. extracted a vocabulary factor and a prior knowledge factor, and then regressed the factors on reading comprehension. When entered last, the vocabulary factor explained a significant 15.1% of the variance in total text recall, but a non-significant 1.1% of the variance in main idea recall. The authors suggest that vocabulary aids sentence-level comprehension (building microstructures), whereas background knowledge aids passage-level comprehension (building macrostructures).

Think-aloud studies. A number of think-aloud protocol studies in reading have identified vocabulary strategies used by high school students. Fehrenbach (1991, see p. 44) identified “failure to understand a word” and “going to another source” (e.g., a dictionary) as vocabulary-related strategies (“going to another source” also included asking someone for help in pronouncing a word, which is a decoding strategy, not a vocabulary strategy).

Kletzien (1991, 1992) found that both good and poor 10th-11th grade comprehenders used “looking for key vocabulary or phrases” (e.g., “I couldn’t do that one because I didn’t know what *hegemony* means” 1991, p. 75 or “I put laws because of constitutional; I know that word” 1992, p. 200) and “paraphrasing,” defined as the student using his or her own words to substitute for words in the text. Each student read texts that were easy (at the independent level), medium-hard (instructional) and difficult (frustration) for that student. In Kletzien (1991), there was no difference in vocabulary strategy use between good and poor readers on independent or instructional level texts.

On frustration level texts, poor readers used more vocabulary strategies than did good readers.

In Kletzien (1992) poor readers used more vocabulary strategies on collection texts but *fewer* vocabulary strategies on causation texts. Both good and poor comprehenders used more vocabulary strategies in frustration-level texts than they did on easier texts, and vocabulary strategies were the most frequently used strategy overall. However, “looking for key vocabulary or phrases” included items that reflected both understanding and lack of understanding, so it is difficult to ascertain the relationship between vocabulary strategies and reading comprehension. Harker (1994) also gives an example of a 10th grade student asking the experimenter for the definition of a word; after being supplied the definition, he was then able to summarize the stanza in the poem he had just read.

Olshavsky (1976-77) collected verbal protocols from 24 tenth-grade good and poor readers. She found “synonym substitution,” “use of context to define a word” (both successful use of context and failure to define), and “stated failure to understand a word.” Vocabulary strategies were the most often used strategies, with “synonym substitution” making up the bulk of vocabulary strategies. Context was used more often by good readers and stated failure to understand a word was used more often by poor readers.

Olshavsky (1978) replicated her study with 12 eleventh-grade good and poor readers using text at 4 levels of difficulty. She found no differences in vocabulary use across reading proficiency.

Correlational studies. Cross-sectional studies conducted since the early 20th century also show a significant relationship between vocabulary and reading comprehension across a wide span of ages. For example, de Jong and van der Leij (2002)

found correlations of .44-.48 in 1st grade; Torgesen, Wagner, Rashotte, Burgess, and Hecht, (1997) found correlations of .49 in 2nd grade and .57 in 3rd grade; Stanovich (1988) found correlations of .50 in 3rd grade, .51 in 5th grade, and .70 in 7th grade; and Thorndike (1973) found correlations of .71 at age 10, .75 at age 14, and .66 at age 17-18 (across 15 countries). Published correlations from studies with students from middle school and older are summarized below in Table 7.

Table 7

Reported Correlations between Vocabulary and Reading Comprehension, by Author

Study	Correlation	Participants	Vocabulary measure
Cain et al. (2004; see p. 61)	.52 for 4 th graders and .63 for 6 th graders	92 4 th grade and 80 6 th grade students from a mixed SES school followed longitudinally, extremely good and extremely poor readers were excluded	Gates-MacGinitie
Coladarci & McIntire (1988)	.71	9,849 12 th grade students	High School and Beyond
Cunningham & Stanovich (1997)	.47	27 11 th grade students (follow up on longitudinal study)	Researcher-developed
Davey (1987)	.42	10 th grade students with low reading comprehension	Gates-MacGinitie
De Soto & De Soto (1983; see p. 71)	.71	134 4 th grade students	Researcher-developed opposites test
Dixon et al. (1988; see p. 72)	.54	95 college undergraduates	Nelson-Denny
Mahony (1994)	.51 to .68	24 9 th grade students	Morpheme Sensitivity Test
McBride-Chang et al. (1993)	.59	49 5 th -9 th grade students	Stanford

Study	Correlation	Participants	Vocabulary measure
Singer & Crouse (1981; see p. 68)	.76	127 6 th grade students	Gates-MacGinitie
Schulman & Havighurst (1947)	.75	97 9 th grade, 82 10 th grade	Seashore-Eckerson vocabulary test
van Gelderen et al. (2004; see p. 62)	.75	281 Dutch 8 th grade students from a range of academic tracks	Researcher-developed
Walczyk (1990)	.78	37 4 th grade students, mostly Caucasian	Iowa Test of Basic Skills
Walczyk & Taylor (1996)	-.41 (latency scores)	109 college undergraduates	Six short expository texts

In summary, there is evidence from experimental studies that vocabulary has a direct effect on reading comprehension. Twenty-five path analysis, regression, think-aloud and correlational studies were also found that are consistent with this effect. Studies were conducted with students in 4th grade through college, reading both narrative and expository texts. These findings hold across standardized and researcher-developed vocabulary measures, as well as measures of morphological knowledge; vocabulary substitution studies; and a variety of vocabulary instruction interventions.

Path 9: Effect of inference on comprehension. Inference is sometimes held not to be separate from comprehension—many studies test literal and inferential comprehension, considering them two aspects of the same phenomenon. However, there is a body of research on 7-8 year old good decoders with poor inference skills and poor comprehension that shows inferencing can be remediated with training (see Oakhill & Yuill, 1996). These results suggests that inference is a separable component of

comprehension. Likewise, strategy instruction research (cf. NRP, 2000) often includes instruction in inferencing. However, psychological theories of reading comprehension have assigned inferencing a critical role, distinct from cognitive and metacognitive strategies, which warrant treating it as a separate variable.

Experimental non-intervention studies. Paris and Lindauer (1976)

compared the ability of children in kindergarten and grades 2 and 4 to draw inferences. Children were presented with test sentences such as, “The truckdriver stirred the coffee in his cup,” and were later asked to recall the sentences. Students who inferred that the driver used a spoon mentioned a spoon in sentence recalls significantly more often than students who did not make such inferences. Older children made significantly more inferences than did younger ones.

Van den Broek et al. (2001) compared delayed free recall performance of 60 students each from 4th, 7th, and 10th grades and college. Participants read 2 stories written at a 3rd grade level, in three inference conditions: questions designed to encourage inferences embedded in the text, questions after the reading, or no questions. Overall recall was significantly greater for the 10th grade and college students than for the younger students. Tenth-grade students who were given questions recalled less than did controls; that is, questions designed to encourage inferences interfered with recall for high school students.

Hare et al. (1989) compared 258 students in 4th, 6th, and 11th grades on ability to identify explicit and implicit (requiring inferences) main ideas in textbook paragraphs. Text structures included listing, sequence, compare and contrast, and cause and effect. Each student read texts at a 3rd grade level and also texts at his or her appropriate grade level. There were significant main effects for grade and interactions between grade and

explicitness at both levels of text. Older students were significantly better able to make the inferences required to identify implicit main ideas than were younger students.

Wagner and Rohwer (1981) compared inferencing in 5th and 11th grade students ($n = 128$) reading unelaborated and elaborated passages. In the elaborated passages, 5th grade students performed as well as the older students. However, in unelaborated passages, 11th grade students were significantly better able to draw inferences than were younger students.

Davey (1988) found that poor 9th-10th grade readers made significantly more errors on reading comprehension questions requiring inference than did good 5th-6th grade readers matched with the older students on reading comprehension.

In two studies with low- and high-skilled college readers, Long, Oppy and Seely (1994; 3 experiments, $N = 600$) and Long et al. (1999; $N = 164$) found similar results. Students read experimenter-developed passages and responded True or False to probe words. Skilled readers showed significant differences in response times to appropriate and inappropriate probes, indicating that they were forming inferences. Less-skilled readers showed no difference at short probe presentation times, but eventually differentiated if given long enough presentation times.

Experimental intervention studies. Inferencing treatment studies are not well known, however, these have been conducted with students in the elementary grades through college. E. Kintsch (1990) created texts that varied in topical organization (macrostructure) and connectedness (microstructure), and tested how well students in 6th and 10th grades and college ($n = 32$ per grade) performed at making inferences in a written summary. Inferences included generalizations, elaborations from prior knowledge, reordering, and connectives or bridging inferences (e.g., use of “in contrast”).

The 6th grade students used significantly fewer of all of the types of inference than did high school and college students (who did not differ). Older students also used significantly more macropropositions and proportionately fewer details in their summaries than did the 6th grade students.

Tenenbaum (1977) compared low- and high-comprehending high school seniors on an inference recognition task. High comprehenders performed significantly better on inference recognition than did low comprehenders, and students who read text with explicit causal links performed significantly better on inference recognition than those who read text without links.

Hannon and Daneman (1998) manipulated texts in order to encourage more reader inferences. Experiment 1 compared 41 skilled undergraduate readers (approximately the upper one-third of 131 college students who took the Nelson-Denny comprehension test) to 38 less-skilled readers (approximately the bottom one-third). Using the passages and probe methodology from Long et al. (1994), Hannon and Daneman replicated their results. There were no differences in latency between a keyword and no-keyword condition. In Experiment two, 35 skilled and 44 less-skilled readers showed the same results. There were no differences in latency between an embedded-question and no-question condition. In Experiment three, 48 less-skilled readers at a slower presentation rate than in Experiment 2. Embedded questions did produce a significant difference in latencies between appropriate and inappropriate probes at this presentation rate. In Experiment four, 23 less-skilled readers completed the same tasks using a repeated-measures design, with the same results. In short, good comprehenders show evidence of making inferences, and embedded questions in text can encourage poor comprehenders to do so, provided enough presentation time is allowed.

Beishuizen, Le Grand, and van der Schalk (1999) gave small-group instruction on inferencing to 60 sixth-grade students, 30 high ability and 30 low ability. Treatment students scored significantly higher on researcher-designed inference questions, but not on a Dutch standardized comprehension test.

Dewitz, Carr and Patberg (1987) compared training in how to complete cloze-type exercises (CL) with a graphic organizer (OV) treatment, a combined treatment (CO), and a control group for 101 5th grade students with a range of reading ability. In an immediate posttest, the CL and GO students answered inference questions equally well, and significantly more than OV and control students. In a 6-week delayed test, CL students performed significantly better than the other three comparison groups.

Path analysis studies. In an SEM study, Britton et al. (1998, see p. 41) found an indirect effect (equivalent to a beta weight) of .30 from inference via background knowledge to reading comprehension.

Singer and Crouse (1981, see p. 68) used a cloze measure with 6th graders and found that inference had a significant effect of .17 on comprehension for 127 6th grade students in a path model that included vocabulary, letter discrimination, non-verbal IQ, and decoding.

Regression studies. Cain et al. (2004; see p. 61) found that scores on a researcher-developed inference test explained a significant 6% of variance in reading comprehension for 4th grade students and 7% of variance for 6th grade students, after accounting for word reading accuracy, vocabulary, IQ, and working memory. Saarnio et al. (1990, see p. 59) found that inference (as measured by a cloze task) explained a significant 6% of variance in reading comprehension for 3rd graders and 8% for 5th

graders when entered last. However, Haenggi and Perfetti (1994, see p. 42) did not find that inferencing made a significant contribution to comprehension for undergraduates.

Think-aloud studies. Nine think-aloud protocol studies with middle and high school students found differences in inferencing across either reader groups or text type. Neuman (1990) compared 21 low-achieving and 21 high-achieving 5th grade students (measured with the Metropolitan Achievement Test) on a think-aloud task with two stories from a children's mystery series. Students' inferences were coded and analyzed; low- and high-achieving readers did not differ significantly on the types or frequency of strategies, but they did differ on inference errors. In making inference errors, low-achieving readers significantly over-relied on their own background knowledge (which sometimes conflicted with the stories), focused more on decoding than on relations between facts, and were not able to impose order on incoherent text.

Phillips (1988) found similar results with 40 low-proficiency and 40 high-proficiency 6th grade students reading 3 passages about either a high-familiarity or a low-familiarity topic. Students thought aloud while reading, but if they did not spontaneously make inferences, probe questions were asked (e.g., "Why was the net hard to pull?" p. 220). Low-proficiency students used ineffective inference strategies (digressing, reverting to a previous interpretation, assuming a default interpretation, and repeating a previous interpretation) significantly more often than did high-proficiency students.

Wilson and Hammill (1982) asked 40 poor, average, good, and superior 9th grade comprehenders (at 25th, 50th, 75th, and 99th percentiles on the Iowa Tests of Educational Development) to summarize verbally while reading aloud from a social studies textbook passage. Poor readers made fewer inferences than expected, and superior readers made more inferences than expected.

Fehrenbach, (1991) found gifted high school students made more inferences than did average students, but only on frustration-level texts, not on independent or instructional level texts. Kletzien, (1992) found that good 10th-11th grade readers made more inferences than did poor readers on a collection (list-type) text, but not on comparison or causation text. Olshavsky (1976-77) found, to the contrary, that poor 9th grade readers used inference more than did good readers. Christopherson, Schultz, & Waern (1981) found high school students made more inferences when reading Bransford and Johnson's (1972) laundry text with a title than when reading the same text without a title. Pritchard (1990) found that 11th grade students used inference more on familiar than on unfamiliar texts.

Earthman (1992) found that literature graduate students thinking aloud while reading poetry more often made inferences, either within the text or from prior knowledge, than did college freshmen.

Correlational studies. Significant correlations between inference and reading comprehension have occasionally been reported in the literature and are summarized in Table 8. There is a slight suggestion of a decreasing relationship with age.

Table 8

Reported Correlations between Inference and Reading Comprehension, By Age

Study	Correlation	Participants	Text
Walczyk (1990)	.65	37 4 th grade, predominantly Caucasian	Three narratives of 275-336 words each
Cain et al.(2004; see p. 61)	.52 4 th grade .61 6 th grade	92 4 th grade and 80 6 th grade students from a mixed SES school followed longitudinally, extremely good and extremely	Short narrative passages

Study	Correlation	Participants	Text
		poor readers were excluded.	
Walczyk & Taylor (1996)	.30 for anaphor reference accuracy -.29 for anaphor reference latency (quick response times associated with high comprehension)	109 college undergraduates	Six short expository texts
Walczyk et al. (2001)	.25	76 college undergraduates	Philosophy text
Dixon et al. (1988; see p. 72)	.16 (NS)	95 college undergraduates	Nelson-Denny
Britton et al. (1998)	.10 (NS) to .28 across measures	211 Air Force recruits age 17-25.	Vietnam War text
Graesser & Bertus (1998)	.32 for younger .61 for older adults	40 younger adults (<i>M</i> age = 22) 40 older adults (<i>M</i> age = 67)	24 five-sentence science passages

In summary, there is evidence from 12 experimental studies that inference has a direct effect on comprehension. Fifteen path analysis, regression, think-aloud, and correlational studies were also found consistent with this effect. Studies tested participants from 2nd grade through older adults. Findings are relevant to several different sub-types of inferencing: within-passage, prior knowledge, and anaphoric reference (determining who a pronoun such as “he” refers to). Treatments included text manipulations as well as training students directly. The findings also hold across several different types of measures, including cloze, multiple-choice, reaction time, and think-aloud protocols.

Path 10: Effect of word reading on vocabulary. A large body of theory (e.g., Perfetti, 1985) and research suggests that readers must be able to decode a word before any stored meaning for that word can be activated. This is the essence of Coltheart's (Coltheart & Rastle, 1994) mediated route in lexical access, as opposed to a direct route from the whole word to its meaning. However, the vast majority of the research has been conducted with children of elementary school age.

Experimental non-intervention studies. One experimental study provides evidence for a direct effect of word reading on vocabulary. White, Graves and Slater (1990) followed 288 students in 1st through 4th grades. Children completed both paper-and-pencil vocabulary measures and also defined the same words in an interview (that is, word reading skills could be partialled out from the latter task). Students who could decode the words were significantly better able to define them orally. However, results were not separated out by grade level.

Path analysis studies. Singer and Crouse (1981; see p. 68) found that decoding loaded a significant .40 on vocabulary for 127 6th grade students in a path model that included vocabulary, letter discrimination, non-verbal IQ, and a cloze measure of inference.

Regression studies. McBride-Chang et al. (1993, see p. 62) found that word identification significantly predicted vocabulary among 49 5th-9th grade students in a regression with age and print exposure also included as predictors ($R^2 = .41$; the correlation between word identification and vocabulary was a significant .62). This study did not separate out results by grade level.

Correlational studies. In a longitudinal study, Wagner et al. (1997) found a steady increase in the correlation between word reading (whether measured by the

Woodcock Word Identification or Word Attack subscale) and vocabulary from kindergarten ($r = .25$) through 4th grade ($r = .60$). This correlation could be due to a direct effect of word reading on vocabulary (or *vice versa*) or could be due to a third factor (e.g., print exposure; Stanovich, 1986). Seven other correlational studies are summarized in Table 9.

Table 9

Reported Correlations between Word Reading and Vocabulary, By Author

Study	Correlation	Participants	Word Reading Measure
Cain et al. (2004; see p. 61)	.57 for 4 th graders .46 for 6 th graders	92 4 th grade and 80 6 th grade students from a mixed SES school followed longitudinally, extremely good and extremely poor readers were excluded	Neale word reading accuracy
Cunningham & Stanovich (1991; see p. 94)	.55	134 students in 3 rd -5 th grades	Accuracy
Cunningham et al. (1990)	-.61 for latency of real word reading -.47 for latency of pseudoword reading	80 undergraduate students	Latency of real word and pseudoword reading
De Soto & De Soto (1983; see p. 71)	.55 for pseudoword accuracy -.51 for pseudoword latency -.48 for real word latency	134 4 th grade students	Latency and accuracy of pseudoword reading; latency of real word reading
Dixon et al. (1988; see p. 72)	.40	95 undergraduate students	Fluency

Study	Correlation	Participants	Word Reading Measure
Jackson & McClelland (1979; see p. 69)	.44 for long passage .55 for short passage	52 undergraduate students	Speed on long and short passages
Singer & Crouse (1981; see p. 68)	.46	127 6 th grade students	Pseudoword accuracy

In summary, there is evidence from one experimental study that word reading has a direct effect on vocabulary. Seven regression and correlational studies were also found that are consistent with this effect. In addition, there is a strong theoretical expectation that word reading has a direct effect on vocabulary (Coltheart & Rastle, 1994; Perfetti, 1985). Finally, word reading has sometimes been found to have a direct effect on vocabulary with younger students (e.g., 2nd grade; Eldredge et al., 1990; but see Aarnoutse & Van Leeuwe, 2000). Due to the conflicting evidence and theory for this path, it was selected as one of the paths to be tested in the variations of the IM model.

Across the ten direct effects hypothesized in the IM model, there is therefore evidence from 130 experimental studies. An additional 123 path analysis, regression, think-aloud and correlational studies were also found that were consistent with these effects. Having summarized the evidence for each effect in the model, I now summarize correlational research that supports hypothesized correlations among the exogenous variables in the model (i.e., background knowledge, vocabulary, and word reading).

Path 11: Correlation between background knowledge and vocabulary. Stanovich (1986) theorizes that because of Matthew effects, students who read a great deal will gain a larger amount of both background knowledge and vocabulary than students who read

less. This should result in a significant correlation between measures of background knowledge and vocabulary.

Cunningham and Stanovich (1991) found a correlation between background knowledge and a vocabulary measure of .43 for 134 students in 3rd-5th grades. Cunningham and Stanovich (1997) found correlations of .55-.84 for 56 eleventh-grade students. Stanovich and Cunningham (1992) found a correlation of .63 for 300 undergraduate students. Stanovich et al. (1995) reported correlations between background knowledge and vocabulary measures of .36-.54 for 133 college students and of .73-.71 for 49 older adults (mean age of 80).

These four studies provide evidence for a correlation between background knowledge and vocabulary; however, no studies were identified that speak to whether there are any direct effects, or whether the correlation is due to a common third factor (e.g., print exposure; see Stanovich, 1986).

Path 12: Correlation between background knowledge and word reading. Due to Stanovich's (1986) Matthew Effects, students who have good word reading skills are also expected to read a great deal, and thereby gain a larger amount of background knowledge than students with poorer word reading skills. This should result in a significant correlation between measures of word reading and background knowledge.

Cunningham and Stanovich (1991) found a significant correlation between background knowledge and word reading accuracy of .27 for 134 students in 3rd-5th grades. Haenggi and Perfetti (1994) found significant correlations of -.19 to -.26 between background knowledge and word reading speed (context-free and in context, respectively) for high school students.

These two studies provide evidence for a correlation between word reading and

background knowledge; however, no studies were identified that speak to whether there are any direct effects, or whether the correlation is due to a common third factor (e.g., print exposure; see Stanovich, 1986).

Summary of the IM model. In summary, the IM model hypothesizes direct effects of background knowledge, inference, strategies, vocabulary, and word reading on reading comprehension. The model further hypothesizes an indirect effect of background knowledge on comprehension via strategies and inference, an indirect effect of strategies on comprehension via inference, an indirect effect of vocabulary on comprehension via inference, and an indirect effect of word reading on comprehension via vocabulary.

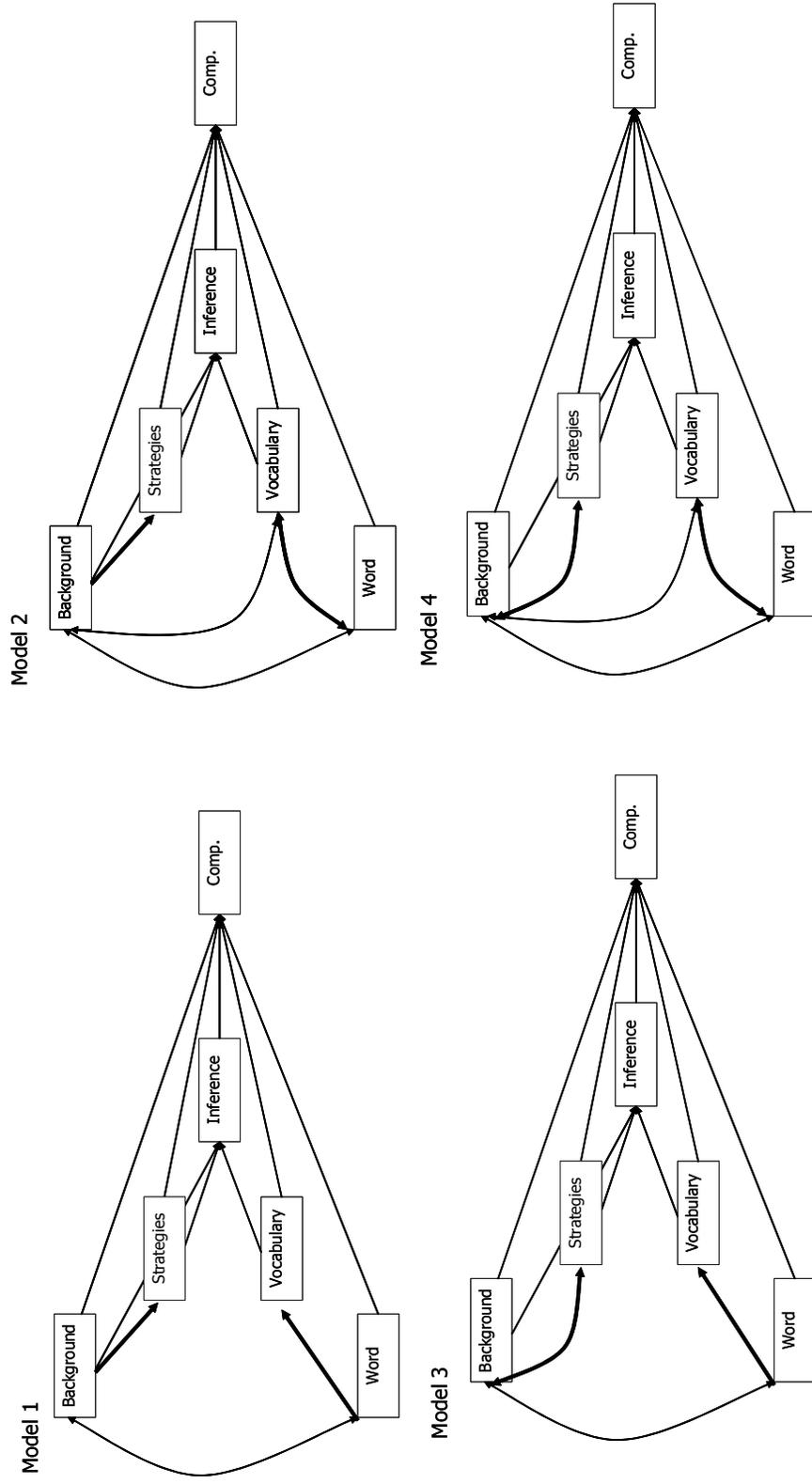
The body of studies reviewed above provides the type of evidence required for a path analysis. First, there is acceptable experimental evidence to establish each of the ten effects in the IM model. Second, for all effects additional path analysis, regression, think-aloud, and/or correlational studies were located consistent with the path. Third, for the two correlations among exogenous variables, correlational evidence was found.

Variations on the IM Model. As discussed above, two paths in the IM Model have conflicting support: the path from background knowledge to strategies (see p. 47) and the path from word reading to vocabulary (see p. 91). There are therefore four models to be tested, shown in Figure 6:

1. Model 1: The IM model plus the WORD→VOC path
2. Model 2: Model 1 without the WORD→VOC path (the same as the IM model)
3. Model 3: Model 1 without the BKGD→ STRAT path
4. Model 4: Model 1 without either the WORD→VOC or BKGD→ STRAT paths

Figure 6

The Four Variations of the Inferential Mediation Model



Having laid out the IM model and its variations, I will now discuss the other “lens” for the current study—measures of the process of reading comprehension using think-aloud methodology. I begin by reviewing previous multi-method (i.e., product and process) studies and some methodological issues associated with them.

Coordinating Product and Process Data

One powerful methodology in the study of both expertise and cognitive development has been the use of coordinating product (e.g., number of balance beam problems solved) and process (e.g., strategy use inferred from verbalizations) data (see Chi, Glaser, & Farr, 1988; Ericsson, 1996; Ericsson & Simon, 1993; Ericsson & Smith, 1991; Feltovich, Ford, & Hoffman, 1997; and Siegler, 1996 for examples). The purpose of this section is to review several recent examples of converging product and process data, considering four themes:

1. The type of product and process data collected
2. How theoretical considerations drive the data analysis
3. Approaches to and quality of data analysis used to coordinate product and process data
4. What the coordination of product and process data adds to our understanding of the phenomena under consideration

The third issue is an important one; given that people vary widely in total number of verbalizations, raw frequency data must be converted to proportions in order to compare participants. Further mathematical conversions may then be necessary, depending on whether a statistical test is conducted, and if so, which one (e.g., an arcsine transformation for ANOVA or regression; median splits for chi-square, etc.). If statistical tests are conducted using raw frequency data or untransformed proportion data, the

results of those statistical tests cannot be interpreted, because assumptions of the tests (e.g., independence of observations, variance independent of the mean) have been violated (Myers & Well, 2003; Pedhazur, 1997).

Another appropriate approach for analyzing process data is to assign participants a scale score, e.g., 0 points for absence of a behavior across the data collection unit, 1 point for some evidence of the behavior, and 2 points for much evidence of the behavior. Inferential statistics may then be appropriately calculated for these data.

Below, six studies in domains other than reading are reviewed, followed by a comprehensive review of multi-method reading studies.

Coordinating Product and Process Data in Science and Mathematics. The studies below add new process data such as gesture, and build on classic multi-method studies in, e.g., chess, physics, Tower of Hanoi, and the Raven's matrices derived from think-aloud data (Chi et al., 1988; Ericsson, 1996; Ericsson & Simon, 1993; Ericsson & Smith, 1991; Feltovich, Ford, & Hoffman, 1997; Siegler, 1996). Think-aloud methodology has been used frequently in cognitive psychology (for reviews, see Ericsson & Simon, 1993 and Newell & Simon, 1972) and in studies of reading comprehension in basic psychology (for a review, see Graesser et al., 1994) and reading research (for a review, see Pressley & Afflerbach, 1995). There is abundant evidence that asking participants to verbalize thinking concurrently does not change cognitive processing, memory, or comprehension provided that participants are not cued to use specific strategies, although thinking aloud does slow down performance, as does reading aloud generally (Ericsson & Simon, 1993; Ericsson & Kintsch, 1995). Think-aloud studies form the foundation of most cognitive strategy instruction interventions, in that flexible strategy use by expert readers, and poor

strategy use by poor readers, were first identified in think-aloud studies (Pressley & Afflerbach, 1995).

In biology, Chung, de Vries, Cheak, Stevens, and Bewley (2002) collected verbal protocol, frequency of using different aids (e.g., a lab simulation, expert opinions), and online log files (e.g., which pages were accessed, and how many times) for 46 undergraduates solving a genetics problem using a hypermedia assessment system. Students also completed multiple-choice measures of prior knowledge and reasoning (scientific, syllogistic, and inferential). The goal of the study was to determine whether the assessments were, in fact, tapping scientific reasoning (i.e., to provide evidence of validity for the measure). Chung et al. appropriately used Spearman rank correlations to test relationships within the proportion of think-aloud and online behaviors, as well as between these two and the multiple-choice measures. They found significant positive correlations between proportion of deep processing in the think-aloud (making accurate cause-effect inferences) and both prior knowledge and scientific and inferential reasoning scores. There were significant negative correlations between proportion of shallow processing in the think-aloud (echoing text, gaps in knowledge, and confusion) and scientific or syllogistic reasoning. High performance (as measured by a large number of problems solved or solving in few trials) was significantly positively associated with deep processing; low performance with shallow processing. Furthermore, high performance was positively associated with use of lab simulations; low performance with use of library and dictionary resources (perhaps reflecting lack of prior knowledge). Here, the process (both verbalizations and computer log files) helped explain *why* students differed in their performance, as well as providing evidence that the assessment was, in fact, measuring scientific reasoning.

In physics, Robertson (1990) asked 15 undergraduate physics students (novices) to think out loud while solving problems related to Newton's second law of motion. Robertson scored the think-aloud data by awarding +1 point for at least one instance of verbalizing an indicator of understanding (e.g., treating the mass of 2 contiguous bodies as separate; there were 5 possible indicators of understanding). He scored -1 point for at least one instance of indicators of misunderstanding in the think-aloud data (e.g., totaling the mass of 2 contiguous bodies; there were 3 possible indicators of understanding). Robertson coordinated product (posttest questions correct) and process (scores derived from the concurrent think-aloud) data by regressing the sum of the indicator scores—which could range from -3 to 5—on posttest performance for similar physics problems. Students who verbalized more indicators of understanding answered significantly more posttest problems correctly. The think-aloud (process) data contributed by explaining *why* the more successful students were able to answer questions correctly.

M. Perry and Elder (1997) compared the concurrent verbalizations and gestures of 6 undergraduates who did advance their understanding (from pretest to posttest) of how gears work to 10 undergraduates who did not. Perry and Elder coordinated process and product data by cross-tabulating verbalization and gesture data. Students who did learn showed more discrepancy between their gestures and verbalizations during learning than did the non-learners. For example, a student might verbalize that a gear on the right was turning clockwise (incorrect), but indicate a counter-clockwise movement (correct) by her hand gesture. The authors hypothesize that verbal-gesture discrepancy is a marker for cognitive conflict that drives learning. Gestures added to the verbal data because they explained why some students with similar verbalizations learned, while others did not.

Hecht (1999) asked 61 undergraduate students to solve arithmetic problems and immediately verbally report the strategies they used. The solution to each of 55 problems was categorized as, e.g., retrieval, decomposition, rule-based, etc. Previous research had suggested that, unlike children who use a range of strategies, adults always use a retrieval strategy. Hecht coordinated product (correct solutions) and process (immediate verbal report) data by computing separate regressions on solution times for problems solved using each different strategy. Retrieval led to significantly faster solution times for participants with higher math SAT scores. Verbal reports added to the solution time data because many students who answered both correctly and incorrectly had similar solution times, but used different strategies. Think-aloud data also revealed that undergraduates use a variety of strategies for solving arithmetic problems, not just retrieval.

Alibali, Bassok, Solomon, Syc, and Goldin-Meadow (1999) coordinated concurrent gesture and verbal report data from 20 undergraduate students as they solved 6 algebra word problems, testing Perry and Elder's (1997) gesture-verbalization discrepancy hypothesis. The word problems involved either continuous units (e.g., air inflating a hot-air balloon), discrete units (e.g., the number of chairs in each row in a lecture hall), or mixed problems that could be described using either continuous or discrete units. Students' verbalizations were coded as continuous, discrete, both, or neither. Students' gestures were then coded as continuous (e.g., sweeping), discrete (e.g., a series of three or more taps), or neutral. Alibali et al. (1999) coordinated the product (correct answer) and process (verbalizations and gesture) data by cross-tabulation. They found that participants were more likely to answer a problem correctly if their gestures reinforced their verbalizations, in support of Perry and Elder's hypothesis. The gesture

(process) data added to the understanding of the verbal reports, because for incorrectly solved problems, the gestures explained *what* the participant's misunderstanding was.

Azevedo, Cromley, and Seibert (2004) used a Self-Regulated Learning (SRL) framework to analyze students' learning about the circulatory system using a hypermedia encyclopedia. They collected paper-and-pencil and verbal protocol data from 51 undergraduates who were randomly assigned to one of three experimental conditions: adaptive scaffolding (AS), fixed scaffolding (FS), and no scaffolding (NS). The AS participants had access to a tutor, while the FS participants had access to 10 researcher-designed questions, and the NS participants learned on their own without any scaffolding provided. In addition to declarative knowledge measures, Azevedo et al. scored essay and diagram measures to examine participants' shift in mental models from pretest to posttest. They then analyzed participants' verbalizations using an SRL coding scheme. They found that mental models for participants in the AS condition increased significantly more than did those in the FS or NS conditions, which did not differ from each other. Chi square analyses of verbal protocol data indicated that participants in the three conditions used different SRL variables. Participants in the AS condition, who learned the most, verbalized significantly more prior knowledge activation, judgments of learning, feeling of knowing, finding location in environment, summarizing, inferencing, help seeking behavior, expressing task difficulty, and expressing interest.

In summary, these six recent studies combined product (e.g., number of questions correct on pre- and post-test) and process (e.g., verbalizations, log files, gestures) data in domains other than reading. They tested specific hypotheses about *how* product and process data would relate to each other (except for Robertson, 1990). Alibali et al. (1999) tested a specific hypothesis about patterns in the process data that were likely to be seen,

based on results from analysis of the product data. The researchers used various appropriate statistical (Spearman rank correlation, regression) and non-statistical (cross-tabulation) data analysis techniques to coordinate the product (scale) and process (frequency) data. In all six cases, the combination of product and process data explained more together than each would have explained alone.

Coordinating Product and Process Data in Reading. Below, I review all reading studies that coordinate product and process data that I was able to identify. I review the studies using the same four criteria as above: types of data collected, theoretical considerations, data analysis, and what is added by using a multi-method design. Reading think-aloud studies that compare the strategies of readers at different proficiencies are, in fact, a multi-method design. However, my discussion here will focus on reading studies that collect think-aloud protocols together with some other form of on-line or process data (e.g., reading time) or product data (e.g., probe reaction time, lexical decision, word naming, Stroop task, or recall data) that can reveal comprehension processes. Overall, the reading studies suffer from a number of data analysis problems that make their results uninterpretable.

Four lines of research have used specific theories to make predictions about how process and product data align in reading. One theory-driven line of multi-component research has emerged from Walczyk's Compensatory-Encoding Model (Walczyk, 1994, 2000; Walczyk & Taylor, 1996; Walczyk et al., 2001). Walczyk and Taylor (1996, see p. 76) tested the prediction that students with slow basic reading processes (i.e., decoding) can compensate with increased metacognitive monitoring and control in the form of text lookbacks. They tested 109 undergraduate students reading 6 short texts on a computer using a moving-window paradigm. They found significant correlations among various

inference, vocabulary, and working memory measures. They also calculated Spearman rank correlations (ostensibly because of skewness in the data) between text lookbacks and the component measures, and found significant correlations with reading time, vocabulary, and working memory measures. Their reported Spearman correlation between text lookbacks and comprehension cannot be interpreted.

Walczyk et al. (2001) collected think-aloud and inference, vocabulary, word reading, and working memory measures from 76 undergraduate students reading a philosophy text and answering literal and inferential questions. The study tested Walczyk's Compensatory-Encoding Model (C-EM), so the think-aloud behaviors of interest were pausing, looking back in the text, and re-reading (see p. 76). Like researchers before them, Walczyk et al. computed Pearson correlations between raw frequency of pauses, lookbacks, and re-reading and scale measures of comprehension, inference, vocabulary, word reading, and working memory. Their reported results are therefore uninterpretable. The studies emerging from the Compensatory-Encoding Model therefore have a strong grounding in theory but have problems with the methods of data analysis.

Another theory-driven line of multi-component research has emerged from Magliano and Graesser's (1991) three-pronged method for studying inferencing. Because of the benefits and limitations of both think-aloud and reaction time process data (and other scale-level data such as probe reaction time, lexical decision, word naming, Stroop task, or recall data), they recommended that studies use three simultaneous approaches. First, discourse processing theories are used to predict specific types of inference that are likely to be used when reading a specific text. Second, think-aloud data are analyzed to identify inferences actually made. Third, behavioral measures such as reaction time or

others listed above are collected to assess whether the predicted types of inferences are actually generated on-line. Long and Bourg (1996) and Whitney and Budd (1996) have also recommended converging think-aloud data with other process measures, though without specifically endorsing the three-pronged method.

Trabasso and Suh (1993) used the three-pronged method with 8 undergraduate students reading eight one-paragraph narratives. First, a discourse processing model was used to identify the inferences that were likely to be made with the texts. Then think-aloud protocols were collected and analyzed for the types of inferences actually made. Patterns in inferencing were then qualitatively compared to priming, reading time, immediate free recall, and coherence rating data from other studies using similar participants reading the same texts, and showed similar patterns. For example, many think-aloud participants might articulate an inference when reading Sentence 1 of a passage, and reaction-time data might show longer time spent in Sentence 1 (relative to sentences in which no inferences were verbalized). This is considered converging evidence from both product (reaction time) and process (think-aloud) data that inferences had, in fact, been drawn. It is not clear, however, whether it is legitimate to compare think-aloud data from one participant to, e.g., priming data from a participant in another study unless both data are from a very large representative samples.

Magliano, Trabasso and Graesser (1999) likewise converged think-aloud data from 48 undergraduates in Experiment 1 with True/False, reading time, and recall data from 76 different undergraduate participants in Experiment 2, using the same 8 short stories. In Experiment 1, participants were instructed to use four particular inference strategies when reading and thinking aloud: understanding, associating, predicting, and explaining. Participants also provided written free recalls 2 days later. Analysis of think-

aloud protocols revealed that the instructions led participants to engage in a large proportion of inferencing, and the mean proportion differed across conditions (.77 to .94). There were no significant differences across conditions in free recall.

In Experiment 2, participants silently read the same passages on a computer screen sentence-by-sentence, wrote answers to understanding, associating, predicting, or explaining questions, answered true/false questions, and provided a free recall. Reading times were also captured by the computer. Participants in the explain condition had significantly longer reading times than participants in the understand condition. There were no significant differences across conditions in true/false questions or free recall. By coordinating these two findings, the authors conclude that participants can be encouraged to use particular strategies, that use of these strategies leads to different patterns in inferencing as shown by both think-aloud data and reading time data. Unfortunately, there were several data analysis problems with this study that make the results uninterpretable: both ANOVAs and regressions were performed between raw frequencies of inferencing and scale data such as idea units recalled and reading time. The studies emerging from the three-pronged method therefore have a strong grounding in theory and use more appropriate methods of data analysis.

McNamara (2003) created a multiple-strategy intervention called Self-Explanation Reading Training (SERT) based on Kintsch's (1988) Construction-Integration model. SERT includes modeling of and prompting self-explanation while reading, as well as direct instruction in monitoring, paraphrasing, predicting, making bridging inferences, using logic, and elaborating. McNamara taught SERT to 21 undergraduates who also thought aloud while they were practicing using newly-taught strategies, and compared them to a think-aloud only control group ($n = 21$). Product

measures included a measure of background knowledge, passage-specific literal and inferential comprehension questions for practice texts and one transfer text, as well as a standardized reading comprehension measure (Nelson-Denny); the process measure was the coded think-aloud protocols. SERT students had significantly higher scores on training and transfer texts; for the transfer text, this difference was entirely due to low-prior knowledge SERT students answering more literal questions correctly. McNamara then analyzed a subset of verbal protocols from 13 SERT students and compared these to 13 control students. She coded for the accurate and inaccurate use of the 6 SERT strategies (monitoring, paraphrasing, predicting, making bridging inferences, using logic, and elaborating) and for rereading. McNamara compared frequency of verbalization of each strategy (corrected for the number of sentences in each passage, but not corrected for the number of verbalizations per participant). She also reported Pearson correlations between the frequency of verbalization of each strategy and scores on the background knowledge, passage-specific comprehension questions, and the Nelson-Denny. In addition, within each self-explanation attempt, she tallied the co-occurrence of different strategies, and analyzed these with chi square tests. Unfortunately, these results cannot be interpreted because the frequency of verbalizations was not corrected for the total number of verbalizations for each participant and Pearson correlations were used for frequency and ratio data.

Using a schema theoretic framework, García (1991) collected retrospective interview data from a subset of 18 5th and 6th grade children. The full sample of 104 students had previously completed measures of both general and text-specific background knowledge, vocabulary, and both literal and inferential comprehension. Fifty-one of the participants were Anglo and 53 were non-ESL bilingual Hispanic children. For the full

sample, Hispanic students scored significantly lower on vocabulary, background knowledge, and comprehension. In a retrospective interview given within one day of taking the comprehension test, students were asked several questions, including explaining how they determined their answers on all three measures. García presents examples from these interviews to illustrate how Hispanic students' lack of background knowledge (e.g., thinking that an antelope could eat a monkey) and vocabulary (e.g., confusing 'an advantage' with 'taking advantage of' led them to choose incorrect answers to comprehension questions.

While these seven studies were theory-driven, four additional exploratory multi-method reading studies were located. Muth, Glynn, Britton and Graves (1988) collected both think-aloud, reading time, and free recall data from 32 undergraduates reading 600-word texts about energy sources and mental illness. Participants were assigned to conditions with and without instructional objectives (questions), and thought aloud while reading versions of the text in which the salient information came early or late in the text. The purpose of the study was to determine why instructional objectives have the effect that they do; that is, when objectives focus students' attention on ideas, what do students do when their attention is focused? Four types of rehearsal during reading were coded: repeating words or sentences, paraphrasing, and summarizing. Recall correlated a significant .29 with reading time. Unfortunately, the authors conducted a Pearson correlation between raw frequency of rehearsal and reading time and free recall data. The results, therefore, cannot be interpreted.

Moore and Scevak (1997) collected think-aloud, written free recall, standardized comprehension, and literal and inferential question answers from 119 of the highest-skilled readers in 5th ($n = 37$), 7th ($n = 40$), and 9th ($n = 42$) grades reading text and either

a diagram (in a science text) or a table (in a history text). Their principal verbal protocol coding categories were main ideas, supporting details, general strategies (including evaluation, imagery, monitoring, prior knowledge activation, rereading, and self-questioning), and using diagrams (including relating diagram to text). They conducted ANOVAs on raw frequency data, which cannot be interpreted. They also conducted cluster analyses on raw frequency data, without controlling for different numbers of verbalizations across participants. They found no significant differences in reading comprehension across the clusters, which is not surprising given that they only selected high-comprehending students for the study. With regard to the free recall and multiple-choice outcomes, they found only 4 significant ANOVAs out of 30 tests. These researchers attempted to coordinate process (i.e., think-aloud) and product (reading comprehension, free recall, and multiple-choice) data using ANOVAs to compare clusters. Unfortunately, there were numerous methodological flaws with this study—they selected only the highest-comprehending students, they used a coding scheme that is very different from that used in prior research, they conducted ANOVAs on the raw think-aloud frequencies, and they used raw frequency counts to define their clusters.

Laing and Kamhi (2002) collected think-aloud, recall, and literal and inferential question answers from 40 3rd grade average and below-average students, as measured by the Woodcock Reading Mastery Test comprehension sub-test. Laing and Kamhi compared students' performance on both reading and listening tasks. The research question was whether the same inference deficits that had been seen in preliminary work with 3rd grade students would appear in both reading and listening conditions. They scored think-aloud protocols for correct and incorrect inferences (including explanations, predictions, and associations). Average and below-average readers differed significantly

on recall and literal and inferential questions. They inappropriately conducted ANOVAs between the proficiency groups using raw frequency of inferential statements, and also computed correlations between raw frequency data, standardized measures, and free recall, making all of their analyses uninterpretable.

Kunz, Drewniak, and Schott (1992) investigated 32 university students with high and low background knowledge reading from a computerized text about meteorology. They collected data on time spent reading, computer traces of students' reading activity (e.g., skipping back a page to re-read), and retrospective cued verbal protocols. The research questions were to explore the on-line self-monitoring processes of college students and to determine whether low-knowledge students could compensate with high levels of self-monitoring. Students read a 3,000 word text on meteorology, then provided a video-cued verbal protocol explaining what they were thinking at 5 selected episodes where they spent longer than usual on a page. Students then answered reading comprehension and near transfer test questions. The reported correlations between scale level data (knowledge level, performance on test questions) and frequency (page skips, mean frequency of strategy use) cannot be interpreted. These four studies overall suffer from both a poor theoretical foundation and poor methods of data analysis.

In summary, then, a few reading researchers have tried to converge reading product (free recall; written open-ended, true/false and multiple choice questions; inference, vocabulary, vocabulary, and working memory) and process data (concurrent think-aloud; retrospective protocol; reading time; page skips/lookbacks). Most of these studies analyzed only a some aspects of all think-aloud verbalizations (e.g., lookbacks), unlike many previous think-aloud studies in reading (see Pressley & Afflerbach, 1995). Some of these studies tested specific theoretical predictions, however in some of them

analyses of data were more exploratory. All of the reading studies suffer from data analysis problems that make most of their results uninterpretable.

Despite these problems with existing reading research, however, there is no reason why multi-method reading studies cannot collect product and process data designed to provide evidence for theoretically-driven hypotheses, using appropriate statistical tests or non-statistical methods. Studies that follow these standards have the potential to add to our understanding of reading comprehension and can point toward appropriate interventions, more so than product or process data can alone. In the case of the current study, direct and indirect paths from background knowledge, inference, strategy use, vocabulary, and word reading to reading comprehension in the IM model (product data) were identified, and then corroborating evidence was sought in the think-aloud protocols (process data).

Statement of the Research Problem

The problem of interest is what struggling adolescent readers in 9th grade struggle with. There is clear evidence *that* many young adolescents struggle with reading, but little evidence of *what* they struggle with, and many competing single-variable hypotheses. In the present study, using the IM model, the direct and indirect effect of each of the 5 predictor variables to reading comprehension was investigated for 9th grade students across a range of reading proficiencies. This analysis can show which variables have the largest effect on comprehension, thereby indicating what components the struggling readers have difficulty with, given significant differences on scores for those variables for low- and high comprehenders.

The question of how each component directly and indirectly affects reading comprehension was investigated through two lenses; path analysis using the component

measures and think-aloud data. Component measures include background knowledge, inference, strategies, vocabulary, and word reading, as well as reading comprehension. A subset of the participants provided a think-aloud protocol while reading from a high school social studies textbook. The think-aloud protocols were then analyzed using an existing coding scheme. The relationships between component and think-aloud data were investigated using two methods: Spearman rank correlations and the co-occurrence of specific pairs of codes in the protocol data.

Expectations. With regard to the fit of variations of the IM model, there are no specific expectations with regard to which variation of the model might fit best. However, it is hypothesized on the basis of the pilot study (Cromley & Azevedo, 2004a) that the original IM model will have an acceptable fit to the data. With regard to the total effect of each of the components, it is expected that vocabulary and inference will make the largest contribution to comprehension, based on results from the pilot study. With regard to differences between low and high comprehenders, it is expected that mean scores for high comprehenders across all five predictors will be significantly higher than those for low comprehenders, but there are no specific expectations with regard to the magnitude of the differences. With regard to the alignment of the path model and inter-relationships among the think-aloud data, it is hypothesized that some paths from the path model will be corroborated by the think-aloud data. It is further expected that the paper-and-pencil measures and the think-aloud data will show slightly different pictures of students' reading comprehension. However, there are no specific expectations of exactly which paths will be corroborated and which ones will not.

Research Questions

The research questions are:

1. Using a new sample, which has the best fit to the data: the CI, VE, or IM model?
2. What is the best-fitting of four related IM models for 9th grade readers?
3. What are the predictor variables that make the largest total contribution to reading comprehension in the best-fitting model for 9th grade readers?
4. How do high- and low-comprehending readers differ on those predictor variables?
5. How are those predictors revealed in the think-aloud protocols of 9th grade readers?

The Four Models To Be Tested

Four models will be tested (see Figure 6 on p. 96):

1. Model 1: The IM model plus the WORD→VOC path
2. Model 2: Model 1 without the WORD→VOC path (the original IM model)
3. Model 3: Model 1 without the BKGD→STRAT path
4. Model 4: Model 1 without either the WORD→VOC or BKGD→STRAT paths

The rationale for these variations on the models is as follows: the IM model contained some paths with contradictory evidence. In the case of the WORD→VOC path, while the path is strongly expected based on theory (e.g., Kintsch, 1998; Perfetti, 1985), only two very weak pieces of evidence for students above 3rd grade were found, and so

the path was not included in the original IM model. Evidence for this path is reviewed above (i.e., White et al., 1990; McBride-Chang et al., 1993).

In the case of the BKGD → STRAT path, as discussed above, the evidence is split on whether background knowledge does have an effect on strategies (see pp. 47-51), with six experimental studies suggesting an effect, and five experimental studies failing to find evidence for an effect.

Four models are therefore proposed; Model 1, including both paths that have conflicting evidence; Models 2 and 3, each with one of those paths; and Model 4 with neither of those paths.

CHAPTER III: PRELIMINARY STUDY

Rationale

A preliminary study was conducted as part of Research Apprenticeship coursework in 2003-2004.¹ The primary goal of this study was to compare the fit of the newly-developed Inferential Mediation Model (referred to as Model 2 in the previous chapter) to that of the Construction-Integration (CI) model (Kintsch, 1988, 1998) and Verbal Efficiency (VE) theory (Perfetti, 1985) with a sample of 9th-grade students across a range of reading abilities.

Method

Participants. Participants were 63 ninth-grade students (28 males and 35 females) from 4 classes in a large high school (serving more than 3,000 students) in a suburb of a large mid-Atlantic city. Participants were 14-15 years old ($M = 14.44$, $SD = .50$). They were diverse both racially (30% White, 34% Black, 25% Hispanic, and 12% Asian) and socio-economically. Five percent of the mothers and 6% of the fathers had not completed high school, whereas 39% of the mothers and 44% of the fathers had an advanced degree. They were also linguistically diverse—64% of students reported their native language as English only, 21% Spanish (alone or with English), 3% Chinese, and 12% other (e.g.,

¹ This research was partially supported by an AERA/Spencer Pre-dissertation Fellowship, funding to Dr. Roger Azevedo from the National Science Foundation (REC#0133346) and the University of Maryland's College of Education and School of Graduate Studies. An earlier version of this material was presented at the annual meeting of the American Educational Research Association in San Diego, CA (April, 2004). I am very grateful to Dr. Azevedo for assisting with design of the study; development of the background knowledge and inference measures; providing the standardized measures, copying, binders, audiotapes, recording equipment; and assisting with data collection on the group and think-aloud measures.

Amharic). These demographics were similar to those of the school as a whole.

Participants were selected to be at a range of reading proficiencies, drawn from two honors (i.e., high GPA) and two non-honors social studies classes. None of the students were in pull-out ESL or Special Education classes.

Research Design. The research design combines maximum likelihood path analysis (Bentler, 1995) using measures of reading comprehension and components (tests of word reading, background knowledge, vocabulary, reading strategies, and inferencing) with think-aloud (Ericsson & Simon, 1993; Pressley & Afflerbach, 1995) and recall data in a multi-method design. Between-subjects factors were comprehension, its components, recall, the think-aloud codes, and the verbal recall score. After separate analyses, product and process data were then coordinated for participants who completed all measures.

Materials and Measures. Written materials included parental consent and student assent forms, and a student background information sheet (See Appendix C); tests of word reading, measures of background knowledge, vocabulary, reading comprehension, reading strategies, and inference; and the think-aloud practice and think-aloud texts. The obtained reliability for all of the measures are reported with the results of the study.

Word reading. The Letter-Word Identification (LWI) and Word Attack (WA) subtests of the Woodcock Diagnostic Reading Battery (WDRB; Woodcock, 1997a), nationally-normed, individually-administered tests, were used to measure participants' word reading skills (for copyright and test disclosure reasons, the measures are not attached). These subtests include real word reading (LWI; 57 questions) and pseudoword reading (WA; 30 questions). They test students' ability to read words aloud, using either sight word knowledge or decoding or morphological (word segment) strategies. Scores from the two subtests were averaged to yield a Basic Reading Skills

cluster score. The LWI and WA subtests are taken from the WJ-R battery. LWI was standardized on a sample of 308 13-year-old students, and the WA on a sample of 236 students. Published reliability (split-half) for the Basic Reading Skills cluster is .926 ($N = 215$) for 13-year-old students (Woodcock, 1997b). Concurrent validity with the Wide Range Achievement Test—Revised (reading) for 17-year-old students is acceptable at $r(24) = .76$ (MacDonald & Cornwall, 1995).

Scores are reported using a WDRB measure called W, which allows comparisons across tests and grades; the maximum possible score is 589 for LWI and 540 for WA. The mean of the two word reading measures was calculated to obtain Woodcock's (1997b) Basic Reading Skills cluster score (WORD; range 376 to 570).

Background knowledge. A researcher-developed, 33-item multiple-choice test was constructed to measure participants' background knowledge about the content referred to in the passages in the Gates-MacGinitie reading comprehension subtest and the think-aloud passage (see below; See Appendix D for the entire background knowledge measure). Nineteen of the questions test content from the Gates-MacGinitie (BKGDRC), and 14 questions test content from the think-aloud passage (BKGDTA); these items were randomized within the measure. Only the background knowledge score related to the Gates-MacGinitie reading comprehension subtest (BKGDRC) was used for the path analysis. The range of possible scores is from 0 to 19.

Vocabulary. The Gates-MacGinitie vocabulary subtest, Level 7/9, Form S (VOCAB; MacGinitie, MacGinitie, Maria, & Dreyer, 2001), a nationally-normed, 45-item multiple-choice test, was used to measure participants' vocabulary knowledge (for copyright and test disclosure reasons, the measure is not attached). The test was standardized using Fall and Spring administrations on a national, representative sample of

4,318 7th grade students, 4,192 8th grade students, and 3,643 9th grade students. The measure's published internal consistency reliability (raw score K-R 20) is high at .90-.92 for 7th-9th grade students, and its Fall-Spring test-retest reliability is high at $r(781) = .88-.90$ (MacGinitie, MacGinitie, Maria, & Dreyer, 2002). No concurrent validity data are given in the *Technical Report*, however the published correlation with the Gates-MacGinitie reading comprehension subtest is $r(12152) = .74-.77$. The range of possible scores is from 0 to 45.

Reading comprehension. The Gates-MacGinitie reading comprehension subtest, Level 7/9, Form S (COMP; MacGinitie et al., 2001), a nationally-normed, 48-item multiple-choice test, was used to measure participants' overall reading comprehension (for copyright and test disclosure reasons, the measure is not attached). This test includes both narrative and expository text (including science and history texts), and features protagonists from a variety of ethnic backgrounds, as well as both male and female characters, similar to participants in the study. The Gates-MacGinitie has both questions that require students to make inferences and literal questions that do not require inferences. The test was standardized on the same population as the vocabulary subtest, above. The measure's published internal consistency reliability (raw score K-R 20) is high at .91-.93 for 7th-9th grade students, and its Fall-Spring test-retest reliability is moderate at $r(781) = .74-.86$ (MacGinitie et al., 2002). No concurrent validity data are given in the *Technical Report*, however the published correlation with the Gates-MacGinitie vocabulary subtest is $r(12152) = .74-.77$. The range of possible scores is from 0 to 48.

Inferences. A researcher-developed 20-item, multiple-choice test was constructed to measure participants' ability to draw inferences using the content referred

to in the passages in the Gates-MacGinitie reading comprehension subtest (INFERRC, 12 items) and the think-aloud passage (INFERTA, 8 items; See Appendix E for the entire measure). Only the inference score related to the Gates-MacGinitie test (INFERRC) was used for the path analysis. The range of possible scores is from 0 to 12.

Reading strategies. The Metacognitive Awareness of Reading Strategies Inventory (STRAT; Mokhtari & Reichard, 2002), a published, 30-item self-report measure using a 5-point Likert-type scale (where 1 stands for “I never or almost never do this” and 5 stands for “I always or almost always do this”), was administered to measure participants’ knowledge of reading strategies (for copyright reasons, the measure is not attached). The measure was validated on sample of 443 students in grades 6-12, including 226 students in grades 7-9. The measure’s published internal consistency reliability (Cronbach’s alpha) is high at .86-.87 for 7th-9th grade students. No concurrent validity data are available, however the developers found significant differences in MARSII scores between different self-reported levels of reading ability (students with lower self-reported reading skills tended to score lower on the MARSII; Mokhtari & Reichard, 2002). The range of possible scores is from 30 to 150.

Think-aloud materials. For the think-aloud portion of the study, students first practiced thinking out loud by reading a short text about Bali (adapted from Roller, 1986; See Appendix F). The passage was re-typed and a red dot embedded in the text after each sentence as a reminder to students to verbalize what they are thinking (see Crain-Thoreson, Lippman, & McClendon-Magnuson, 1997, for a discussion of the minimal impact of verbalization reminders). The passage is 99 words long and has a Flesch-Kincaid grade level of 8.0. It was printed single-spaced in 12-point type on one 8.5 x 11 sheet of paper.

After thinking aloud about the practice passage, participants produced a think-aloud protocol while reading a passage about the Revolutionary War taken from a high school social studies textbook (Viola, Wheatley, & Hart, 1998; See Appendix G). The passage was scanned in color from a textbook and a red dot was embedded in the text after each sentence. The passage was 1,025 words long and had a Flesch-Kincaid grade level of 9.2. It included 2 maps and one illustration with a caption, including an 18th-century cartoon. The passage was printed on 3 separate pages, which were placed in front of the participant. Participants had access to a pen and paper with which to take notes (see, e.g., Wade et al., 1990), but were not permitted to write on the text.

Equipment. All individual sessions (word reading and think-aloud) were audiotaped on a cassette recorder using a clip-on microphone. Students in the think-aloud session were provided with pen and paper so that they would have the opportunity to take notes. The think-aloud sessions was transcribed from the audiotapes using a Sony transcribing tape player.

Procedure. A cover letter and parental permission forms were sent home, and teachers collected the signed forms. After obtaining parental permission, student informed assent was given (see Appendix C), and all students completed the previously mentioned background information sheet. All students then completed the component measures in the following order: individually-administered word reading; group-administered background knowledge, vocabulary, reading comprehension, inference, and strategy use. Fourteen of the participants then completed a think-aloud protocol in a separate session beginning the day after component measures were completed, and continuing until all protocols were collected.

Component measures. Students completed the word reading measures (LWI and WA), in an individual session with me in a private office at the school lasting approximately 10 minutes. Instructions for administering the tests are as follows: the test book with lists of words (LWI) or pseudowords (WA) is set up on the table in front of the participant, who is asked to read each word aloud in turn. The researcher turns the pages until all words on the page are errors, or the end of the subtest has been reached. No feedback or corrections are given. For the LWI test, published instructions are designed to minimize participant frustration and testing time, while producing reliable results (Woodcock, 1997b). For the LWI, participants began with Item 30 (the suggested starting point for students in grades 5-9), and continued reading each page of 6 items until all words on the page were errors (ceiling method). In the event that the student read all of the first 6 items incorrectly (basal method), the examiner is instructed to turn the pages backward one at a time and have the student read each page until all items on the page are read *correctly*. For the WA test, all participants began with Question 1 and continued until an entire page of pseudowords were read incorrectly, or the end of the subtest was reached. Sessions were tape-recorded, and correct and incorrect answers were recorded while administering both subtests.

In their regular classroom, each class then completed the following group-administered measures in the following order: background knowledge, vocabulary, reading comprehension, inference, and strategy use, in one 90 min session (one class period in a block schedule). For each test, participants were told the amount of time allowed, the number of questions, and where to write their answers. Participants were asked to do their best, and to give their best answer if they were not sure. Participants were also asked not to work on the other tests if they finished early. Time was allowed

for participant questions. Participants whose parents had not given permission or who did not themselves assent to participate were asked to read a set of readings copied from the source materials used to develop the Gates-MacGinitie reading comprehension subtest while the rest of the class completed the measures. Researchers remained in the classroom during the entire test period to answer questions, ensure students were recording their answers on answer sheets and remained on task, and ensure the security of test materials.

The background knowledge test was administered in an 11-minute whole-class session to each class in their regular classroom. After completing one practice question, participants were instructed to read the question and mark the single best answer for each of the 33 questions on a separate answer sheet.

The vocabulary test was then administered in a 20-minute whole-class session to each class in their regular classroom, following the published instructions for administering the test. After completing two practice questions, participants were instructed to mark the single best answer for each of the 45 questions on a separate answer sheet.

The reading comprehension test was then administered in a 35-minute whole-class session to each class in their regular classroom, following the published instructions for administering the test. After completing two practice questions, participants were instructed to mark the single best answer for each of the 48 questions on a separate answer sheet.

The inference test was then administered in a 7-minute whole-class session to each class in their regular classroom. After completing one practice question, participants

were instructed to read the passages and questions and mark the single best answer for each of the 20 questions on a separate answer sheet.

The strategy use instrument was then administered in a 10-minute whole-class session to each class in their regular classroom. After completing one practice question, participants were instructed to read each question and circle the number that applies to them on the sheet for each of the 30 questions.

Think-aloud. Fourteen students at a range of reading comprehension proficiencies (based on their Gates-MacGinitie scores) were then selected to complete the think-aloud portion of the study. None of the participants had prior experience with think-alouds, to my knowledge.

The think-aloud session was then conducted individually in an office at the school, and the entire session was tape-recorded. Each student practiced thinking aloud and produced a think-aloud protocol during a session lasting approximately 30 minutes.

Each student first practiced thinking aloud by reading the practice text about Bali, which took less than 5 min. During training no feedback was given, however, participants were prompted to think out loud until they produced at least 3 verbalizations (see Crain-Thoreson et al., 1997). During training participants were reminded to think aloud, if necessary, with one of three reminders: “Please say what you are thinking,” “Don’t forget to read out loud,” or “What are you looking for now?” We did not otherwise intervene during the practice, even if students expressed word difficulty, mispronounced or misread words, or asked questions.

After practice, the following directions to participants were displayed and read out loud: “You are being presented with a passage from a high school social studies textbook.

We are interested in learning about how students learn from what they read. I want you to read this passage as if you were learning the material for a class. You have a pen and paper to take notes, if that is what you would usually do when you are studying by yourself, but I will collect them when you are done reading. In order to understand how you learn from a textbook, I need you to think out loud while you are reading. Please say everything you are thinking out loud while you read the text. I'll be here in case anything goes wrong with the tape recorder, but I can't answer any questions about the reading or help you with it. Please remember that it is very important to say everything you are thinking while you are working on this task." These instructions remained visible during the session.

Participants were then given the Revolutionary War text and thought out loud while reading it (approximately 20 min). During the session participants were reminded to think aloud using the same reminders mentioned above. As in the practice session, we did not otherwise intervene during the think-aloud.

Data Analysis and Scoring

The component measures. All component measures were scored and entered into an SPSS file. Measures were scored as follows: LWI and WA were scored in accordance with the WDRB instructions (Woodcock, 1997b). For LWI, the participant was given 1 point each for the first 29 questions (the "floor" level), and 1 point each for all subsequent correct answers. Words had to be pronounced conventionally; correct phonetic decoding was not considered correct and no partial credit was given (e.g., if the term "deja vu" were given, it would have to have been pronounced with the French pronunciation, not "dee-jah vuh"). The W score, mentioned previously, was then recorded from the appropriate norm table. For WA, 1 point was given for each correct

answer; no partial credit was given. The W score was then recorded from the appropriate norm table. For both LWI and WA, student answers were checked while students were being tested and then re-checked later from the audiotape.

For the background knowledge and inference measures each correct multiple-choice answer received one point. Ambiguous answers (e.g., more than one circle filled in) were counted as wrong answers.

The vocabulary and reading comprehension measures were scored in accordance with the Gates-MacGinitie instructions. One point was given for each correct answer, and scores were then recorded from the appropriate norm table. Ambiguous answers (e.g., more than one circle filled in) were counted as wrong answers.

For the inference measure, each correct multiple-choice answer received one point. Ambiguous answers (e.g., more than one circle filled in) were counted as wrong answers.

For the strategy use measure, the numerical answer (1-5) for each item was entered in an SPSS file. Ambiguous answers (e.g., more than one number circled) were counted as missing data.

Think-aloud data. Each think-aloud session was transcribed from the audiotape, segmented, and later coded (see below). Each tape was transcribed verbatim, following the conventions of Bracewell and Bruleux (1994). The transcript was segmented into clauses (each containing a subject and a verb; see Alibali et al., 1999 and Magliano et al., 1999). This resulted in a total of 87 typed pages ($M = 6.2$ pages per participant), 20,678 words ($M = 1,477$ words per participant), and 735 coded segments.

Coding scheme for the think-aloud protocols. The coding scheme for the think-aloud protocols was adapted from several SRL frameworks (especially

Pintrich, 2000; Winne, 2001; Winne & Hadwin, 1998) and previous research (Azevedo, Cromley, & Seibert, 2004; Azevedo, Guthrie, & Seibert, 2004; Fehrenbach, 1991; Zwaan & Brown, 1996). In addition, codes were added as they emerged from the data. Major categories for the coding scheme were *background knowledge*, *inference*, *strategies*, *vocabulary*, and *word reading*. Each major category was then subdivided into codes indicating accurate use (e.g., accurately summarizing, indicated by SUM+) and those indicating inaccurate use (e.g., inaccurately summarizing, indicated by SUM- see McNamara, 2003). The codes and definitions, and examples from this study are found in Appendix H.

Background knowledge includes three codes: *Prior Knowledge Activation* (used accurately and inaccurately) and *Anachronisms* (inaccurate). *Prior Knowledge Activation* (PKA) is verbalizing background knowledge (including knowledge learned earlier in this text) relevant to the think-aloud passage. For accurate PKA, historical information is recalled that is both correct and relevant, e.g., “molasses is like syrup I think” (all examples are drawn from the preliminary study; reading from text is shown with underlining). In inaccurate PKA, historical events are recalled but the details are inaccurate “the African-American bus boycott when they refused to buy- to go on the buses to hurt the economy.” *Anachronisms* are statements about events in the past, but ones that could not have happened in the past because of political or other changes, e.g., “people in like, like a conference room trying to get the Stamp Act and Sugar Act away,” and are therefore coded as inaccurate.

Inferencing includes six codes: *Back to the Future*, *Evaluation*, *Hypothesizing*, *Inferences*, *Knowledge Elaboration*, and *Links*. *Back to the Future* was coded when students imagined what they would do if they had been alive in the past, e.g. “He held out

longer than I would have,” and was coded as accurate. *Evaluation* is any moral judgment about what is happening in the text, e.g. “they’re getting way too out of hand,” and is always coded as accurate. *Hypothesizing* is any hypothesis or prediction about events to follow in the text, e.g., “But who should pay to support them? I’m thinking us, not Britain,” and is always coded as accurate. *Inferences* are text-to-text logical deductions, and may be accurate, e.g., “They’re trying to keep their land safe from the Indians I guess” or inaccurate, “Britain would soon feel the effects of that spirit of freedom in the colonies. So I guess he did a lot.” *Knowledge Elaborations* are logical deductions between background knowledge and text, and may be made accurately, e.g., “the British needed big- a huge army” or inaccurately, “So the Stamp Act affected more people like lawyers and newspaper publishers because they’re the ones who need to, uh, mail more stuff and send papers.” *Links* are a specific type of knowledge elaboration in which the participant makes an inference between prior historical knowledge or current events and what was read in the text, and are accurate, e.g., “I am thinking that this reminds me of . . . what we’re learning in history now, which was, or a while ago, which was when they would break into buildings when they thought that someone was communist.”

Strategies includes nine codes: *Feeling of Knowing*, *Imagery*, *Judgment of Learning*, *Not Thinking*, *Re-reading*, *Self-Questioning*, and *Summarizing*. *Feeling of Knowing* is acknowledgment of comprehension, e.g., “that makes sense,” and was coded as accurate. *Judgment of Learning*, by contrast, is an acknowledgement of a lack of understanding or a breakdown in comprehension, e.g., “I’m kind of confused,” and was coded as inaccurate. *Imagery* is stating a mental image of the situation, and is coded as accurate, e.g., “I’m thinking of like soldiers walking” or inaccurate, “Thinking of jury and Englishmen talking, you know, trying.” *Not Thinking* is an explicit statement from

the participant the he or she is not thinking about what was read, and was coded as inaccurate. *Re-reading* is reading out loud for a second (or further) time a passage that was read previously, e.g., “To enforce the Proclamation of 1763 . . . I have to read it over again. To enforce the Proclamation of 1763,” and was coded as accurate. *Self-Questioning* is generating a specific question, e.g., “Now I’m wondering what do their dresses look like,” and was coded as accurate. *Summarizing* includes paraphrasing or restating information from the text in the reader’s own words, and may be accurate, e.g., “So George Grenville, the British Prime Minister, was looking for other ways for them to pay for their own defense” or inaccurate, e.g., “So the Stamp Act had been around for a while but they weren’t really enforced.”

Vocabulary includes two codes: *Vocabulary Difficulty* and *Vocabulary Knowledge*. *Vocabulary Difficulty* is an expression of not knowing the meaning of a word or not being able to figure out the meaning of a word, e.g., “Rivalries, what are those?” *Vocabulary Knowledge* is a paraphrase of a word meaning or use of synonym, e.g., “I guess customs duties were the, uh, were like taxes.”

Word Reading includes two codes: *Word Pronouncing Difficulty* and *Omissions*. *Word Pronouncing Difficulty* is inability to pronounce a word conventionally (even if the pronunciation is phonetically accurate), e.g., “Col-o-nel” for Colonial; only miscues that might affect meaning were coded. *Omissions* are words that the participant neither read nor re-read, e.g., “There should be no [New] Englanders”; only omissions that might affect meaning were coded.

Interrater Reliability. I coded the entire corpus, and three transcripts (21% of the corpus) were recoded by a master’s student in educational psychology, who was trained

using the coding scheme and already-coded transcripts. We agreed on 145 out of 161 segments, or 90% of segments. All disagreements were then resolved by discussion.

Results

Data were first screened for ceiling and floor effects; many of the participants had difficulty completing the inference measure in the time allotted. Only data from the first 14 questions were therefore analyzed; the 8 questions related to the content of the Gates-MacGinitie were used in the path analysis.

The strategy use measure showed poor convergent validity with the reading comprehension measure. A subset of the 8 most-highly discriminating questions was therefore used in the path analysis.

Reading Components—Descriptive Analyses. Means and standard deviations for reading comprehension, background knowledge, inference, strategies, vocabulary, and word reading are shown in Table 10.

Table 10

Descriptive Statistics for Reading Comprehension and Component Measures

Component	<i>M</i>	<i>SD</i>
1. COMP (Raw score; Max = 48)	32.6	11.7
2. COMP (percentile)	48.6	33.6
3. COMPLIT (Max = 28)	19.1	7.3
4. COMPINF (Max = 20)	13.5	4.7
5. BKDDRC (Max = 19)	13.5	3.1
6. INFRC (Max = 8)	5.0	2.0
7. STRAT (Max = 40)	24.0	5.3
8. VOCAB (Raw score; Max = 45)	32.7	9.1
9. VOCAB (Percentile)	64.0	28.5
10. LWI (Raw score; Max = 57)	50.8	3.0
11. LWI (W score; Max = 589)	532.4	12.2
12. WA (Raw score; Max = 30)	22.5	4.0
13. WA (W score; Max = 540)	508.2	9.8
14. WORD (W score; Max = 564.5)	520.5	10.1

Note: COMP = Gates-MacGinitie reading comprehension; COMPLIT = literal questions from the Gates-MacGinitie reading comprehension subtest; COMPINF = inference questions from the Gates-MacGinitie; BKGDRC = questions from BKGD related to the Gates-MacGinitie comprehension subtest; INFRC = questions from INF related to Gates-MacGinitie comprehension; STRAT = Metacognitive Awareness of Reading Strategies Inventory (MARS), VOCAB = Gates-MacGinitie vocabulary test; LWI = Letter-Word Identification subtest of the Woodcock Diagnostic Reading Battery (WDRB); WA = Word Attack subtest of the WDRB; WORD = mean of W scores on LWI and WA.

Participants were split at the median on comprehension scores, and low- and high-comprehending students were compared on each of the components. With regard to background knowledge, high comprehenders scored 20% higher than low comprehenders ($t [63] = 3.33, p < .05$). For inferencing, high comprehenders scored 86% higher than low

comprehenders ($t [63] = 5.12, p < .05$). With regard to strategies, high comprehenders scored 31% higher than low comprehenders ($t [63] = 5.94, p < .05$). On the vocabulary measure, high comprehenders scored 40% higher than low comprehenders ($t [63] = 5.96, p < .05$). On the word reading measure, high comprehenders scored a small but statistically significant 2% higher than low comprehenders ($t [63] = 3.48, p < .05$). These components interacted to yield reading comprehension scores for high comprehenders that were 86% higher than those of low comprehenders ($t [63] = 12.16, p < .05$).

Path Analysis. Maximum Likelihood (ML) estimation in EQS (Bentler, 1995) was used to conduct a reliability-adjusted path analysis (Hancock, 1997) on the effect of the 5 reading components (background knowledge, inference, strategies, vocabulary, and word reading) on reading comprehension, using the three models presented above. In reliability-adjusted path analysis, an error estimate ($1 - \text{Cronbach's alpha reliability}$) is entered into the path model, rather than calculating ML estimates of the error terms. Coefficients (path loadings) in the final model are analogous to standardized beta weights in regression. Because the model is reliability adjusted, the path loadings therefore reflect the influence of each variable, after adjusting for the reliabilities of the measures. In these analyses, six latent variables are postulated (one for each predictor variable and one for reading comprehension), and each measured variable was hypothesized to be made up of the latent variable and our estimate of measurement error. Correlations and Cronbach's alpha reliabilities for the observed data are shown in Table 11. The variance/covariance matrix used in the path analysis is shown in Table 12.

Table 11

Correlations and Reliabilities for Reading Measures

Subscale	1	2	3	4	5	6	7	8	9	10	11	12
1. COMP	.93											
2. COMPLIT	.98***	.90										
3. COMPINF	.96***	.88**	.84									
4. BKDDRC	.53***	.49***	.55***	.76								
5. INFERC	.71***	.68***	.70***	.52***	.76							
6. STRAT	.59***	.58***	.56***	.33**	.53***	.71						
7. VOCAB	.72***	.70***	.69***	.74***	.69***	.57***	.93					
8. LWIRAW	.27	.22	.35*	.44**	.43**	.30*	.67***	.74				
9. WARAW	.28	.24	.32*	.51***	.40**	.25	.56***	.72***	.76			
10. LWIW	.30*	.24	.37**	.45**	.45***	.30*	.67***	.99***	.70***	.74 ^a		
11. WAW	.29*	.25	.33*	.50***	.43**	.25	.56***	.70***	.98***	.69***	.76 ^a	
12. WORD	.36*	.30*	.42**	.52***	.52***	.33*	.69***	.93***	.90***	.94***	.90***	.85 ^a

* $p < .05$ (.28 < $r < .36$) ** $p < .01$ (.35 < $r < .46$) *** $p < .001$ ($r > .45$)

^a Reliability for the scaled scores is the same as that of the raw scores. Reliability for the WORD composite is reliability of all 87 LWI and WA items together.

Note: Cronbach's alpha reliabilities are shown on the diagonal and correlations are shown below the diagonal. See Table 10 for abbreviations.

Table 12

*Variance/Covariance Matrix for Reading Comprehension and Predictor Component**Measures*

Subscale	1	2	3	4	5	6
1. BKGDRC	9.16					
2. INFRC	5.78	26.17				
3. STRAT	17.96	23.95	107.32			
4. VOCAB	19.79	26.36	66.19	77.14		
5. WORD	2.94	5.29	9.56	10.97	3.95	
6. COMP	23.30	41.34	67.20	85.37	17.81	155.39

Note: Variances are shown in the diagonal and covariances are shown below the diagonal. See Table 10 for abbreviations.

The fit of the three models to the data was tested using four indices: χ^2/df , AIC (Akaike, 1987), and the two fit indices recommended by Hu and Bentler (1999) for samples of less than 250 participants: CFI and SRMR. Hu and Bentler recommend $CFI \geq .96$ and $SRMR \leq .10$ as indicating acceptable fit. The results of these tests of fit are presented in Table 13.

Table 13

Indicators of Fit for the Three Reliability-Adjusted Models

Fit Index	<i>CI Model</i>	<i>VE Model</i>	<i>IM Model</i>
AIC	1.98	14.31	1.86
χ^2 (df)	15.98* (7)	28.31*** (7)	7.86 (3)
SRMR	.08	.07	.04
CFI	.96	.91	.98
RMSEA (90% CI)	.15 (.05, .24)	.22 (.14, .31)	.162 (.01, .30)
R ²	.89	.91	.90

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

Note: The best-fitting model according to each fit index is indicated in ***bold italics***.

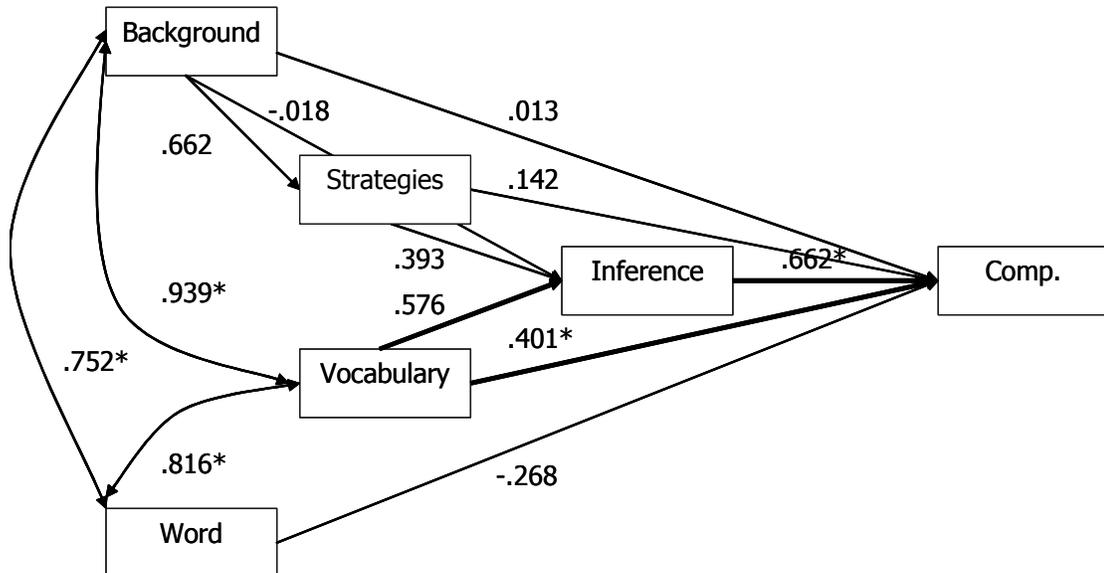
Overall, the IM model had the best fit, as indicated by the smallest χ^2 (which was also non-significant, indicating excellent fit), AIC, and SRMR, and the largest CFI.² The RMSEA is large, but this is probably due to the small sample size. Results of the path analysis for the IM model, including standardized maximum likelihood coefficients (equivalent to beta weights), are shown in Figure 7.³

² Because the models are not nested in each other, there is no statistical test for of comparing the fit of the three models.

³ If the model is run without the reliability adjustment, the IM model still has a much better fit than the CI or VE models (AIC = 11.76, 22.03 and 54.65, respectively), but none of the models meet Hu and Bentler's (1999) criteria for good fit.

Figure 7

Final Solution for the Inferential Mediation Model



For the Inferential Mediation model, two components made a significant, unique contribution to reading comprehension once all other components were controlled for.

The greatest overall contribution to comprehension was made by vocabulary, followed by inference, strategies, background knowledge, and word reading; direct and indirect paths for all three models are summarized in Table 14.⁴

⁴ If the models are run without reliability adjustment, vocabulary still makes the largest contribution to comprehension, followed by strategies, inference, background knowledge, and word reading. In the non-adjusted model, the direct effects of strategies, vocabulary and inference are statistically significant by a one-tailed z tests at $p < .05$, and all indirect and total effects are significant by the same criterion.

Table 14

Direct and Indirect Paths from Predictors to Reading Comprehension, for the CI, VE, and IM Models

Variable	CI Model			VE Model			IM Model		
	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
BKGD	—	-.033	-.033	—	.059	.059	.013	.254	.266
INFER	.809*	—	.809	.956*	—	.956	.662*	—	.662
STRAT	.247*	.006	.253	—	.376*	.376	.142	.260	.402
VOCAB	—	.677*	.677	—	.530*	.530	.401*	.381	.782
WORD	—	.565*	.565	—	.470*	.470	-.268	—	-.268
R ²			.892			.914			.901

* $p < .05$

Note: Dashes indicate path was not in the model

Think-Aloud Data. To compare the use of various components across levels of reading proficiency, participants' raw frequencies of verbalizations were converted to proportions; for example, participant NM used *Summarizing* 8 times out of 39 utterances,

for a proportion of 21%. These proportions were then compared for each variable across proficiency groups. Results of the analyses are presented in Table 15.

Table 15

Proportion of Use for Each Code in the Think-Aloud Coding Scheme, Across Groups

Code	<i>Low-Comprehending</i>	<i>High-Comprehending</i>
	<i>Students</i>	<i>Students</i>
Background knowledge	2.7%	4.9%
Accurate	.4%	3.1%
PKA+	.4%	3.1%
Inaccurate	2.3%	1.8%
PKA-	1.2%	1.8%
ANACH-	1.2%	0%
Inferences	18.8%	38.8%
Accurate	15.2%	34.4%
EVAL+	4.4%	15.3%
HYP+	1.2%	3.1%
INF+	5.8%	9.8%
KE+	2.0%	3.4%
LINK+	1.8%	2.9%
Inaccurate	3.6%	4.4%
INF-	3.6%	2.9%
KE-	0%	1.1%
LINK-	0%	.4%
Strategies	38.7%	37.2%

Code	<i>Low-Comprehending</i>	<i>High-Comprehending</i>
	<i>Students</i>	<i>Students</i>
Accurate	16.2%	19.0%
BTF+	.6%	.4%
FOK+	0%	1.8%
IMAGE+	1.2%	0%
RR+	1.6%	4.1%
SQ+	2.2%	1.8%
SUM+	10.6%	11.0%
Inaccurate	22.5%	18.2%
JOL-	6.2%	7.6%
IMAGE-	2.8%	0%
NOTHINK-	4.4%	2.0%
SUM-	9.0%	8.4%
Vocabulary	7.4%	3.3%
Accurate		
VOC+	.4%	1.5%
Inaccurate		
VOC-	7.0%	1.8%
Word reading		
Inaccurate	30.4%	15.1%
WORD-	26.8%	13.8%
OMIT-	3.6%	1.3%

Note: A + indicates accurate use of the component; a – indicates inaccurate use.

Both low- and high-comprehending students enacted all of the components of comprehension; even the low-comprehending students often tried to enact strategies and make inferences. For the low comprehenders, 38.7% of verbalizations were strategies and 18.8% of their verbalizations were inferences (compared to 37.2% and 38.8%, respectively, for high comprehenders). However, overall, the low-comprehending students were more inaccurate when they attempted to do this, with the biggest differences in inferencing. Across all codes, accurate codes accounted for 58% of verbalizations for high comprehenders but only 42% for low comprehenders.

Results of the Spearman rank correlation analyses on the think-aloud data.

Spearman rank correlations were performed on the proportion of verbalization of each accurate and inaccurate code across participants for the five predictor variables. This analysis results in a correlation corresponding to each direct effect or correlation in the path analysis. The analysis therefore does not provide evidence of a causal relationship, it simply suggests that parallel relationships may exist in the paper-and-pencil and think-aloud data. The correlations are shown in Table 16.

Table 16

*Spearman Rank Correlations Among Proportions of Verbalization for the Five Predictor**Variables*

Code	1	2	3	4	5	6	7	8
1. BKGD+	—			<i>N</i> = 14				
2. INF+	.14	—						
3. STRAT+	.24	-.06	—					
4. VOC+	.50*	-.09	NA	—				
5. BKGD-	NA	.19	-.21	-.41	—			
6. INF-	-.21	NA	-.11	-.16	.14	—		
7. STRAT-	-.30	-.57**	NA	NA	NA	.08	—	
8. VOC-	-.13	-.55**	NA	NA	-.50*	-.22	NA	—
9. WORD-	-.37	NA	NA	-.58**	.19	NA	.14	-.12

* $p < .10$ ** $p < .05$

Notes: NA indicates a relationship not hypothesized in the model. Significant correlations are indicated in ***bold italics***.

Despite limited statistical power, there were significant negative correlations between inaccurate strategy use and accurate inferencing ($r_s [14] = -.57, p < .05$), and between lack of vocabulary knowledge and accurate inferencing ($r_s [14] = -.55, p < .05$). That is, using a high proportion of errorful strategies is associated with a low proportion of accurate inferencing. Expressing much difficulty with vocabulary words is also associated with a low proportion of accurate inferencing.

There were also significant negative correlations between accurate use of vocabulary with word reading difficulty ($r_s [14] = -.58, p < .05$). That is, participants who

expressed understanding of vocabulary also tended to make significantly fewer consequential word reading errors. Overall, high comprehenders tended to simply be more accurate in their use of the components, not to use a different range or balance of strategies compared to low comprehenders.

Discussion

Of the three models, the IM model had the best fit; its fit was somewhat better than that of the CI model, and much better than that of the VE model. One major difference in the models is whether the predictors are allowed to have a direct effect on comprehension. In all three models, inferencing had a large, direct effect on comprehension, and background knowledge had its largest effect indirectly. In the IM model, background knowledge made its contribution via its effect on strategies, the subsequent effect of strategies on inferencing, and the subsequent effect of inferencing on comprehension.

In the IM model, vocabulary and strategies made important contributions to comprehension both directly and indirectly. Vocabulary had an indirect effect on comprehension via inferencing. That is, knowing the meaning of a word allowed readers to draw inferences necessary to comprehend the text. Strategies likewise made their contribution via inferencing; being able to, e.g., use boldface to understand what ideas are important enabled readers to make the logical connections needed to understand what they were reading. Simple regression models that have been used in prior research are not able to measure these important indirect effects.

Word reading accuracy had a non-significant but negative direct effect in the IM model, contrary to prior research and theory. There are two possible explanations. Recall that there were very small differences in word reading accuracy between high and low

comprehenders. It is possible that word reading accuracy does not contribute to 9th-grade students' comprehension, but we noticed large differences in fluency during the think-alouds. Perhaps, as Artelt et al. (2001) suggested, it is word reading fluency rather than accuracy that is the issue for 9th-grade students. A second explanation, given that the CI and VE models did show large significant indirect effects of word reading accuracy, is that word reading accuracy makes an important contribution through its effect on vocabulary, a path that was not included in the IM model.

The results suggest that vocabulary and inferencing make the largest contribution to reading contribution in the model, and that none of the students were at ceiling on these two measures. At the same time, students' profiles were relatively flat—students who struggled with comprehension not only tended to have low inferencing and vocabulary scores, but also low scores on the background knowledge, strategies, and word reading measures.

Implications for the Dissertation Study

This study provided preliminary support for a new model of reading comprehension that includes both direct and indirect effects of background knowledge, strategies, vocabulary, and word reading, and confirms the strong effects of vocabulary and inferencing on comprehension that have been found in many previous studies.

The preliminary study had certain limitations in terms of the sample size, measures, and procedures; these limitations were addressed in the dissertation study. The sample size in the preliminary study was relatively small, and many paths and Spearman rank correlations were non-significant. This led to increasing the sample size for the dissertation study in order to increase power. Participants did not have sufficient time in the preliminary study to complete the inference measure. This led to increasing the

amount of time given for the inference measure in the dissertation study. The strategy use measure in the preliminary study had poor concurrent validity with the comprehension measure. This led to developing a different strategy use measure for the dissertation study, one which includes both Gates-MacGinitie-related items and Revolutionary War items. With regard to the vocabulary measure, the Dissertation Committee pointed out that it was a general vocabulary measure, and there was not a section specific to the vocabulary of the think-aloud passage. This led to modifying the vocabulary measure for the dissertation study. Finally, the odd negative loading for word reading accuracy in the preliminary study led to two changes in the dissertation study. First, a measure of word reading fluency was added to the dissertation study. Second, in the dissertation study, two variations on the IM model were added to test the effect of word reading on vocabulary. In addition, every item on every measure was checked to see if deletion would increase the reliability of the measure, and for the dissertation study items showing little discrimination were deleted.

CHAPTER IV: METHOD

Participants

Participants were 177 ninth-grade students, selected to be at a wide range of reading comprehension proficiency.¹ They were drawn from 9 social studies classes—4 honors, 4 regular, and 1 remedial class—at a large high school (more than 3,000 students) located in the suburbs of Washington, DC. None of the students were in pull-out ESL or Special Education classes. Twenty-one percent of students at the high school receive free or reduced-price meals. Demographics for the whole sample and the think-aloud sub-sample are shown in Table 17.

Table 17

Demographics for the Whole Sample and the Think-Aloud Sub-Sample

	<i>Entire sample</i> (<i>N</i> = 177)	<i>Think-aloud</i> sub-sample (<i>n</i> = 44)
Age		
Mean	14.20	14.09
SD	.55	.56

¹ A sample size of 180 was suggested by an *a priori* power analysis using the data from the preliminary study. Briefly, the model from the preliminary study was re-run as if the sample size had been larger but the observed variance/covariance matrix remained the same. Different sample sizes were entered incrementally until there was a gap in the number of additional paths that were statistically significant.

	<i>Entire sample</i> (<i>N</i> = 177)	<i>Think-aloud</i> sub-sample (<i>n</i> = 44)
Sex		
Male		
<i>N</i>	99	20
%	56%	46%
Female		
<i>N</i>	78	24
%	44%	55%
Race		
White	28%	11%
Black	27%	34%
Hispanic	23%	18%
Asian	14%	25%
Mixed race	8%	11%

Think-Aloud Sub-Sample. The think-aloud sub-sample ($n = 44$) was slightly but non-significantly younger than those not selected to provide think-alouds ($t [71.24] = 1.56, p > .05$). Girls were slightly but non-significantly over-represented in the think-aloud sub-sample compared to those not selected ($\chi^2 [1, N = 44] = 2.61, p > .05$). The distribution of students of different races in the think-aloud sub-sample was non-significantly different from those not selected ($\chi^2 [4, N = 44] = 2.99, p > .05$).

Research Design

The research design combines maximum likelihood path analysis (Bentler, 1995) using measures of reading comprehension and components (tests of word reading, background knowledge, vocabulary, reading strategies, and inferencing) with think-aloud (Ericsson & Simon, 1993; Pressley & Afflerbach, 1995) and recall data in a multi-method design. Between-subjects factors were comprehension, its components, recall, the think-aloud codes, and the verbal recall score. After separate analyses, product and process data were then coordinated for participants who completed all measures.

Materials and Measures

Written materials included parental consent and student assent forms, and a student background information sheet (See Appendix I); tests of word reading, measures of background knowledge, vocabulary, reading comprehension, reading strategies, and inference; and the think-aloud practice and think-aloud texts.

Word reading. Two measures were used for word reading accuracy and one for word reading fluency.

Word reading accuracy. The Letter-Word Identification (LWI) and Word Attack (WA) subtests of the Woodcock Diagnostic Reading Battery (WDRB; Woodcock, 1997a), nationally-normed, individually-administered tests, were used to measure participants' word reading accuracy (for copyright and test disclosure reasons, the measures are not attached). These subtests include real word reading (LWI; 57 questions) and pseudoword reading (WA; 30 questions). They test students' ability to read words aloud, using either sight word knowledge or decoding or morphological (word segment) strategies. Scores from the two subtests were averaged to yield a Basic Reading Skills cluster score. The LWI and WA subtests were taken from the WJ-R battery. LWI was

standardized on a sample of 308 thirteen-year-old students, and the WA on a sample of 236 students. Published reliability (split-half) for the Basic Reading Skills cluster is .926 ($N = 215$) for 13-year-old students (Woodcock, 1997b). Concurrent validity with the Wide Range Achievement Test—Revised (reading) for 17-year-old students is acceptable at $r(24) = .76$ (MacDonald & Cornwall, 1995). In a pilot study with 63 ninth-grade students, the measures showed internal consistency reliability Cronbach's alphas of .74 for LWI and .76 for WA (Cromley & Azevedo, 2004a).

Scores are reported using a WDRB measure called W, which allows comparisons across tests and grades; the maximum possible score is 589 for LWI and 540 for WA. The mean of the two word reading measures is calculated to obtain Woodcock's (1997b) Basic Reading Skills cluster score (WORD; range 376 to 570).

Word reading fluency. Participants read a short passage out loud, and the number of words correctly read in one minute was counted. The 240-word passage, concerning precursors to World War I, was taken from a high-school level social studies passage in the *Qualitative Reading Inventory-III* (Leslie & Caldwell, 2000; see Appendix J for the entire passage). Participants were asked to "Please read the passage out loud as accurately as you can at your normal reading speed." Participants read the first two sentences of the passage, a timer was started unobtrusively, and I continued timing for one minute. The number of total words read and number of errors and omissions during that one minute was then recorded. The possible score ranges from 0 to 240 words read correctly per minute and from 0 to 240 errors.

Background knowledge. A researcher-developed, 20-item multiple-choice test was adapted from Cromley and Azevedo (2004a) to measure participants' background knowledge about the content referred to in the passages in the Gates-MacGinitie reading

comprehension subtest and the think-aloud passage (see below; See Appendix K for the entire background knowledge measure). Thirteen of the questions test content from the Gates-MacGinitie (BKGDRC), and 7 questions test content from the think-aloud passage (BKGDTA); these items were randomized within the measure. Only the background knowledge score related to the Gates-MacGinitie reading comprehension subtest (BKGDRC) was used for the path analysis. The range of possible scores is from 0 to 13. In a pilot study with 63 ninth-grade students using a 20-item measure, the internal consistency reliability Cronbach's alpha was .76 (Cromley & Azevedo, 2004a).

Vocabulary. Participants' vocabulary knowledge was measured with a 34-item measure. Twenty-three of the questions were odd items from the Gates-MacGinitie vocabulary subtest, Level 7/9, Form S (VOCABRC; MacGinitie et al., 2001), a nationally-normed multiple-choice test. I constructed the remaining 11 questions with assistance from Dr. Azevedo to test key vocabulary from the think-aloud passage (VOCABTA). (See Appendix L for researcher-developed items from the vocabulary measure; for copyright and test disclosure reasons, the questions from the Gates-MacGinitie vocabulary subtest are not attached). The Gates-MacGinitie vocabulary subtest items were standardized using Fall and Spring administrations on a national, representative sample of 4,318 seventh-grade students, 4,192 eighth-grade students, and 3,643 ninth-grade students. The published internal consistency reliability (raw score K-R 20) for the entire Gates-MacGinitie vocabulary subtest is high at .90-.92 for 7th-9th grade students, and its Fall-Spring test-retest reliability is high at $r(781) = .88-.90$ (MacGinitie et al., 2002). In a pilot study with 63 ninth-grade students who completed the entire measure, the internal consistency reliability Cronbach's alpha for all items was .93 and for odd items only was .84 (Cromley & Azevedo, 2004a). No concurrent validity data are

given in the *Technical Report*, however the published correlation with the Gates-MacGinitie reading comprehension subtest is $r(12152) = .74-.77$. In the pilot study, the correlation of the entire Gates-MacGinitie vocabulary subtest with the Gates-MacGinitie reading comprehension subtest was $r(63) = .72$ (Cromley & Azevedo, 2004a). The range of possible scores is from 0 to 34.

Reading comprehension. The Gates-MacGinitie reading comprehension subtest, Level 7/9, Form S (COMP; MacGinitie et al., 2001), a nationally-normed, 48-item multiple-choice test, was used to measure participants' overall reading comprehension (for copyright and test disclosure reasons, the measure is not attached). This test includes both narrative and expository text (including science and history texts), and features protagonists from a variety of ethnic backgrounds, as well as both male and female characters, similar to participants in the study. The Gates-MacGinitie has both questions that require students to make inferences and literal questions that do not require inferences. The test was standardized on the same population as the vocabulary subtest, above. The measure's published internal consistency reliability (raw score K-R 20) is high at .91-.93 for 7th-9th grade students, and its Fall-Spring test-retest reliability is moderate at $r(781) = .74-.86$ (MacGinitie et al., 2002). In the pilot study, the internal consistency reliability Cronbach's alpha was .93 (Cromley & Azevedo, 2004a). The range of possible scores is from 0 to 48.

Inferences. A researcher-developed 16-item, multiple-choice test was adapted from Cromley and Azevedo (2004a) to measure participants' ability to draw inferences using the content referred to in the passages in the Gates-MacGinitie reading comprehension subtest (INFERRC, 10 items) and the think-aloud passage (INFERTA, 6 items; See Appendix M for the entire measure). Only the inference score related to the

Gates-MacGinitie test (INFERRC) was used for the path analysis. The range of possible scores is from 0 to 10. In the pilot study, a previous 12-item version of this measure had a Cronbach's alpha internal consistency reliability of .76 and concurrent validity with the subset of inferential questions on the Gates-MacGinitie reading comprehension subtest was $r(63) = .70$ (Cromley & Azevedo, 2004a).

Reading strategies. I constructed a 16-item, multiple-choice test with assistance from Dr. Azevedo to measure participants' ability to apply specific strategies (i.e., summarizing, searching, activating prior knowledge, and self-questioning) to the same passages used in the *Inference* measure described above (STRATRC, 10 items) and the think-aloud passage (STRATTA, 6 items; See Appendix M for the entire measure). This measure was modeled on one developed by Kozminsky and Kozminsky (2001). Only the strategy score related to the Gates-MacGinitie test (STRATRC) was used for the path analysis.²

The background knowledge, vocabulary, reading comprehension, reading strategies, and inferencing measures, together with an answer sheet, were placed in 3-ring binders with tabs for ease of administration.

Think-aloud materials. The think-aloud session included two measures: a think-aloud protocol and an immediate verbal summarization protocol. For the think-aloud portion of the study, students first practiced thinking out loud by reading a short text about Bali (adapted from Roller, 1986; See Appendix N). The passage was re-typed, a red dot embedded in the text after each sentence as a reminder to students to verbalize what they are thinking (see Crain et al., 1997, for a discussion of the minimal impact of

² The self-report strategy-use measure originally proposed was completed by 175 students, and had a correlation of .10 with reading comprehension. This measure was therefore dropped in favor of the multiple-choice measure described above.

verbalization reminders), and facts in the passage were updated (i.e., the population of Java was revised from 60 million to 123 million, so the population density was doubled). The passage is 99 words long and has a Flesch-Kincaid grade level of 8.0. It was printed single-spaced in 12-point type on one 8.5 x 11 sheet of paper.

After thinking aloud about the practice passage, participants produced a think-aloud protocol while reading a passage about the Revolutionary War taken from a high school social studies textbook (Viola et al., 1998; See Appendix G). This was the same passage used in the preliminary study. Participants had access to a pen and paper with which to take notes (see, e.g., Wade et al., 1990), but were not permitted to write on the text.

Equipment

All individual sessions (word reading, think-aloud, and verbal recall) were audiotaped on a cassette recorder using a clip-on microphone. A stopwatch was used to time the fluency passage, the think-aloud practice session, the think-aloud protocol, and the verbal recall protocol. Students in the think-aloud session were provided with pen and paper so that they would have the opportunity to take notes. The think-aloud sessions were transcribed from the audiotapes using a Sony transcribing tape player.

Procedure

A letter to parents and the parental permission form was sent home with students, and those students who returned them were asked to give student informed assent and complete the background information sheet (See Appendix I). All of the students who returned parental permission forms and who assented to participate then completed the component measures in the following order: individually-administered word reading; group-administered background knowledge, vocabulary, reading comprehension,

inference, and strategy use. A sub-sample of students ($n = 44$) then completed the think-aloud and verbal recall protocols in a separate session beginning about one month after the component measures were completed (in order to allow time for scoring the measures and in order to select a representative sub-sample).

Component measures. Students completed the word reading measures (LWI, WA, and fluency), in an individual session with me in a private office at the school lasting approximately 10 minutes. The procedures were identical to those used in the preliminary study. For the fluency passage, participants were asked to “read this passage out loud as accurately as you can at your normal reading speed.” I started the timer after participants had read the first two sentences, and marked any errors or omissions on a record sheet while the participant read.

In their regular classroom, each class then completed the following group-administered measures in the following order: background knowledge, vocabulary, reading comprehension, and inference/strategy use, in one session lasting approximately 90 min. For each test, participants were told the amount of time allowed, the number of questions, and where to write their answers. Participants were asked to do their best, and to give their best answer if they were not sure. Participants were also asked not to work on the other tests if they finished early. Time was allowed for participant questions. Participants whose parents had not given permission or who themselves did not assent to participate were offered a set of readings copied from the source materials used to develop the Gates-MacGinitie reading comprehension subtest while the rest of the class completed the measures. I remained in the classroom during the entire test period to answer questions, ensure students were recording their answers on answer sheets and remain on task, and ensure the security of test materials.

The background knowledge test was administered in a 7-minute whole-class session to each class in their regular classroom. After completing one practice question, participants were instructed to read the question and mark the single best answer for each of the 20 questions on a separate answer sheet.

The vocabulary test was then administered in a 15-minute whole-class session to each class in their regular classroom. After completing one practice question, participants were instructed to mark the single best answer for each of the 34 questions on a separate answer sheet.

The reading comprehension test was then administered in a 35-minute whole-class session to each class in their regular classroom, following the published instructions for administering the test. After completing two practice questions, participants were instructed to mark the single best answer for each of the 48 questions on a separate answer sheet.

The inference/strategy test was then administered in a 17-minute whole-class session to each class in their regular classroom. Participants were instructed to read the passages and questions and mark the single best answer for each of the 32 questions on a separate answer sheet.

Think-aloud. Forty-four students selected to be representative of participants as a whole on the Gates-MacGinitie comprehension subtest were then selected to complete the think-aloud portion of the study. They were selected to represent the middle range of comprehension (students with extremely low and extremely low scores were excluded) and to have the same mean comprehension score as the full sample Gates-MacGinitie Mean = 34.0, $SD = 10.1$, Min = 7, Max = 45). Participants for the think-aloud sessions were also screened to ensure that they did not have widely varying profiles across the

component measures (e.g., positive Z-scores on vocabulary and inference but negative Z-scores on background knowledge and word reading, an uncommon pattern in this data set). None of the participants had prior experience with think-alouds, to my knowledge.

The think-aloud session was then conducted individually in an office at the school, and the entire session was tape-recorded. Each student practiced thinking aloud on the practice text (see Appendix N), produced a think-aloud protocol on the Revolutionary War text (see Appendix G), and then produced a verbal recall, during a session lasting approximately 25 minutes.

Each student first practiced thinking aloud by reading the practice text about Bali (see Appendix N), which took less than 5 min. During training no feedback was given, however, participants were prompted to think out loud until they produced at least 3 verbalizations (see Crain-Thoreson et al., 1997). During training participants were reminded to think aloud, if necessary, with one of three reminders: “Please say what you are thinking,” “Don’t forget to read out loud,” or “Say what you are looking for now.” I did not otherwise intervene during the practice, even if students express word difficulty, mispronounced or misread words, or asked questions.

After the practice session, the following directions to participants were displayed and read out loud: “You are being presented with a passage from a high school social studies textbook. We are interested in learning about how students learn from what they read. I want you to read this passage as if you were learning the material for a class. You have a pen and paper to take notes, if that is what you would usually do when you are studying by yourself, but I will collect them when you are done reading. In order to understand how you learn from a textbook, I need you to think out loud while you are reading. Please say everything you are thinking out loud while you read the text. I’ll be

here in case anything goes wrong with the tape recorder, but I can't answer any questions about the reading or help you with it. Please remember that it is very important to say everything you are thinking while you are working on this task." These instructions remained visible on the table during the session.

Participants were then given the Revolutionary War text (see Appendix G) and asked to think out loud while reading it (approximately 20 min). During the session participants were reminded to think aloud using the same reminders mentioned above. As in the practice session, I did not otherwise intervene during the think-aloud.

After finishing the reading, the instructions, passage, and any notes taken were removed, and students were asked to verbally recall information from the text, with the instruction, "Please tell me everything you can remember about what you just read." When they finished, participants were then prompted with the question, "Anything else?" If they added any more statements, they were prompted once more with the same prompt. After any further responses, the session was concluded. The verbal recalls took less than 5 minutes.

Data Analysis and Scoring

Component measures. All component measures were scored and entered into an SPSS file. Measures were scored as follows: LWI and WA were scored in accordance with the WDRB instructions (Woodcock, 1997b). For LWI, the participant was given 1 point each for the first 29 questions (the "floor" level), and 1 point each for all subsequent correct answers. In accordance with the instructions, these real words had to be pronounced conventionally. No partial credit was given, and unconventional but phonetically correct decoding was not scored as a correct answer (e.g., if the term "deja vu" had been given, it would have to have been pronounced with the French

pronunciation, not “dee-jah vuh”). Regionalisms and dialectical variations were allowed. The *W* score, mentioned previously, was then recorded from the appropriate norm table. For WA, 1 point was given for each correct answer; no partial credit was given. For both LWI and WA, student answers were checked while students were being tested and then re-checked later from the audiotape if necessary.

For the fluency measure, both raw words per minute and the number of errors were entered into an SPSS file, and the number of correct words per minute was also calculated (see Jenkins et al., 2003 for a discussion of the use of measures of words per minute versus correct words per minute in fluency research).

For the background knowledge and inference measures each correct multiple-choice answer received one point. Ambiguous answers (e.g., more than one circle filled in) and missing answers were counted as wrong answers.

The reading comprehension measure was scored in accordance with the Gates-MacGinitie instructions. One point was given for each correct answer, and scores were then recorded from the appropriate norm table. Ambiguous answers (e.g., more than one circle filled in) and missing answers were counted as wrong answers.

For the vocabulary, inference, and strategy use measures, each correct multiple-choice answer received one point. Ambiguous answers (e.g., more than one circle filled in) and missing answers were counted as wrong answers.

Principal component scores for word reading. After checking the word reading data for normality, a single score for word reading was calculated for each participant based on the first principal component (PC) extracted from four word reading variables: the Letter Word Identification and Word Attack raw scores, the number of words read correctly while reading the one-minute reading passage, and the number of

errors/omissions (number of words) in the one minute passage. The first PC had an eigenvalue of 2.62 and explained 62.5% of the variance in the four word variables; the second PC had a much smaller eigenvalue of .71 and explained only 17.7% of the variance, so it was not retained. The choice of one PC was confirmed using Velicer's Minimum Average Partial procedure; average squared correlations were at a minimum of .12 with one PC. Loadings on the first principal component were large and in the expected direction (LWI: .91; and WA: .87; Correct words per minute: .79; Errors: -.64). Principal component scores were then calculated from these loadings and saved in SPSS. These principal component scores were then used in the path analyses.

Think-aloud and verbal recall data. Each think-aloud session was transcribed from the audiotape, segmented, and later coded (see below). Each tape was transcribed verbatim, following the conventions of Bracewell and Bruleux (1994). This resulted in a total of 225 typed pages ($M = 5.11$ pages per participant) and 62,065 words ($M = 1411$ words per participant). The transcript was segmented into clauses (see Alibali et al., 1999 and Magliano et al., 1999). I then coded all of the transcripts using the coding scheme described below.

Coding scheme for the think-aloud protocols. The coding scheme for the think-aloud protocols was adapted from the Self-Regulated Learning coding scheme developed by Azevedo and colleagues (Azevedo & Cromley, 2004; Azevedo, Cromley, & Seibert, 2004; Azevedo, Guthrie, & Seibert, 2004), modified based on previous think-aloud reading studies (Fehrenbach, 1991; Laing & Kamhi, 2002; McNamara, 2001; Neuman, 1990; Robertson, 1990; Zwaan & Brown, 1996), and codes that emerged from the pilot study. Major categories for the coding scheme are *background knowledge*, *inferencing*, *strategy use*, *vocabulary*, and *word reading*. Each major category was then

subdivided into codes indicating accurate use (e.g., accurately summarizing, indicated by SUM+) and those indicating inaccurate use (e.g., inaccurately summarizing, indicated by SUM-). The codes and definitions, with examples from the current study are found in Appendix O.

Background knowledge includes three codes: *Prior Knowledge Activation* (used accurately and inaccurately) and *Anachronisms* (inaccurate). *Prior Knowledge Activation* (PKA) is verbalizing background knowledge (including knowledge learned earlier in this text) relevant to the think-aloud passage. For accurate PKA, historical information is recalled that is both correct and relevant, e.g., “That reminds me of the Boston Tea Party,” (all examples are drawn from the current study; reading from text is shown with underlining). In inaccurate PKA, historical events are recalled but the details are inaccurate and/or irrelevant, “[they] said that during the slave era, ‘no taxation without representation’,” or the participant states that he or she lacks background knowledge, e.g., “I don’t know who that person is.” *Anachronisms* are statements about events in the past, but ones that could not have happened in the past because of political or other changes, e.g., “Doesn’t that like go against the Constitution?” when referring to events of the 1760s (the Constitution was ratified in 1788), and are therefore coded as inaccurate. To ensure that verbalizations coded as inaccurate PKA and anachronisms were in fact such, the verbal protocol text and a list of all of these verbalizations was sent to Ms. Kelly Ryan, a doctoral candidate in History at the University of Maryland College Park and a Colonial specialist, who verified the historical facts that were referred to in the verbal protocols. We agreed on 30 out of the 32 references to historical facts that I had coded as inaccurate were in fact inaccurate. The remaining two verbalizations were recoded as correct.

Inferencing includes six codes: *Back to the Future*, *Evaluation*, *Hypothesizing*, *Inferences*, *Knowledge Elaboration*, and *Links*. *Back to the Future* was coded when students imagined what they would do if they had been alive in the past, e.g. “I would have done the same thing,” and was coded as accurate. *Evaluation* is any moral judgment about what is happening in the text, e.g. “I think that’s right” and is always coded as accurate. *Hypothesizing* is any hypothesis or prediction about events to follow in the text, e.g., “They probably wouldn’t like that,” and is always coded as accurate. *Inferences* are text-to-text logical deductions, and may be accurate, e.g., “Because [the British] had a lot of goods to sell” or inaccurate, “They owed money because they lost.” *Knowledge Elaborations* are logical deductions between background knowledge and text, and may be made accurately, e.g., “taking their money against their will. [That’s] illegal” or inaccurately, “sounds like the Indians were very bad.” *Links* are a specific type of knowledge elaboration in which the participant makes an inference between prior historical knowledge or current events and what was read in the text, and may be made accurately, e.g., recalling the television program “Law and Order” when reading about search warrants or inaccurately, “they would be considered guilty unless proven innocent. I think that’s the way the law system works here as well.”

Strategies includes nine codes: *Coordinating Information Sources*, *Feeling of Knowing*, *Imagery*, *Judgment of Learning*, *Not Thinking*, *Re-reading*, *Self-Questioning*, *Summarizing* and *Taking notes*. *Coordinating Information Sources* includes putting a graphic together with the text, e.g., “That picture looks like the stamps down there,” and was coded as accurate. *Feeling of Knowing* is acknowledgment of comprehension, e.g., “that makes sense,” and was coded as accurate. *Judgment of Learning*, by contrast, is an acknowledgement of a lack of understanding or a breakdown in comprehension, e.g., “I

don't understand that," and was coded as inaccurate. *Imagery* is stating a mental image of the situation, and is coded as accurate, e.g., "I'm thinking that, like, British judges look mean." *Not Thinking* is an explicit statement from the participant that he or she is not thinking about what was read, and was coded as inaccurate. *Re-reading* is reading out loud for a second (or further) time a passage that was read previously, e.g., "They planted by your care? No! They planted by your care?," and was coded as accurate. *Self-Questioning* is generating a specific question that the reader expects the forthcoming text to answer, e.g., "I'm wondering what colonies they were," and was coded as accurate (note that a *self-question* need not include a question word; see Graesser & Person, 1994; Scardamalia & Bereiter, 1992). *Summarizing* includes paraphrasing or restating information from the text in the reader's own words, and may be accurate, e.g., "I think he wants them to be united" or inaccurate, e.g., "They tried to sell and buy things with Britain." *Taking notes* was coded for taking written notes with the pen and paper provided.

Vocabulary includes two codes: *Vocabulary Difficulty* and *Vocabulary Knowledge*. *Vocabulary Difficulty* is an expression of not knowing the meaning of a word or not being able to figure out the meaning of a word, e.g., "one thing that was smuggled was— that was stolen or— was molasses." *Vocabulary Knowledge* is a paraphrase of a word meaning or use of synonym, e.g., for "bribing and smuggling," "So apparently there was some corruption" or an explicit statement that the meaning of the word is known.

Word Reading includes three codes: *Self-Correction*, *Word Pronouncing Difficulty* and *Omissions*. *Self-Correction* is when the student mis-pronounces a word, and then corrects him or herself and pronounces it correctly, e.g., "miles acres the sea, across the sea." *Word Pronouncing Difficulty* is inability to pronounce a word

conventionally (even if the pronunciation is phonetically accurate), e.g., “burnded” for burdened; only miscues that might affect meaning were coded. *Omissions* are words that the participant neither read nor re-read, e.g., “There should be no [New] Englanders”; only omissions that might affect meaning were coded.

Verbal recall protocols. The verbal recall protocols were transcribed according to the same conventions as for the think-aloud protocols (see p. 157). Each protocol took up less than one typed page; there was a total of 7,126 words ($M = 162$ words per participant). The major topics of the text (e.g., the Stamp Act; there were 4 major topics), sub-topics ($n = 11$), and supporting facts for each ($n = 50$) were identified *a priori*. Recalls were then scored using a rubric that accounts for the number of major topics (4 points each), sub-topics (2 points each), and supporting evidence (1 point each; see Appendix P for the rubric). The number of errors in the recall (e.g., “in America”) was then subtracted from the score; self-corrections were not counted as errors. Each main idea or supporting fact could only be counted once, and prior knowledge not in the text as well as inferences or knowledge elaboration were not counted. Possible scores therefore ranged from 88 (every major topic, sub-topic and supporting detail stated, and no errors) to -65 (every major topic, sub-topic and supporting detail stated in error, and no correct statements).

Path analyses. To compare the fit of four variations on the IM model to the data, I used EQS (Bentler, 1995) to conduct maximum likelihood path analyses. ML path analysis is similar to multiple regression-based path analysis, except that maximum likelihood estimation is used to calculate loadings (analogous to beta weights) instead of the ordinary least-squares estimation used in regression. ML path analysis has similar

assumptions to multiple regression, namely, independence of observations, multivariate normality, and no multicollinearity (Bentler, 1995).

Coordinating Product and Process Data. To coordinate the findings from the path analysis and the results of the think-aloud protocols, Spearman rank correlations were computed among the predictor variables in the model for the proportion of accurate and inaccurate verbalizations in the think-aloud protocols. In this analysis, the correlations quantify the interrelationships among the think-aloud variables in parallel with the correlations that underlie the path analysis. Participants who provide think-alouds should have scores on the paper-and-pencil measures used in the path analysis that mirror the distribution of scores for all participants. For the purpose of the Spearman rank correlations, the distributions of scores of the think-aloud participants should not vary widely from those for all participants.

Inter-rater Agreement. For the think-aloud protocols, after transcription I coded all 44 think-aloud protocols. No new codes were added to the coding scheme; the only code in the original coding scheme that was not used was IMAGE-. Then, 36% of the corpus (16 randomly-selected transcripts) was recoded by a second coder, who is a graduate student in educational psychology. This graduate student had been trained in the summer of 2003 and had assisted with coding the pilot think-aloud protocols. The coder was re-trained using the definitions and examples given below, and using already-coded segments that were not included in the portion of the corpus to be re-coded. He then re-coded the 16 transcripts. We agreed on 817 out of 870 codes, yielding an interrater agreement of 94%. After re-coding, all differences were resolved by discussion. For the verbal recall protocols, after I scored all of the verbal recalls, 34% of the recalls (15 randomly-selected transcripts) were recoded by a second coder, a high school social

studies teacher and recent Ph.D. in Curriculum and Instruction. The coder first reviewed the rubric (see Appendix P) and did not suggest any changes based on either his knowledge of American History or his teaching experience. He was then trained using the rubric, and using four already-coded recalls that were not included in the portion to be re-coded. We agreed on the scores for 16 out of the 16 protocols, yielding an interrater agreement of 100%.

CHAPTER V: RESULTS

Descriptive Analyses

Reading components measures. Means, standard deviations, skewness, and kurtosis statistics for reading comprehension, background knowledge, inference, strategies, vocabulary, and word reading across the full sample and the sub-sample of participants who provided think-aloud protocols are shown in Table 18. Two participants were missing data for LWI and WA; these cases were deleted, leaving a final sample of 175. As planned, participants spanned the range of reading comprehension levels, from 1st to 99th percentiles, with a mean of 58th percentile. Almost all variables were slightly negatively skewed, indicating that high-skilled students were slightly over-sampled. However, there was no evidence of ceiling effects for any of the measures. Skewness scores with an absolute value < 3 and kurtosis scores with an absolute value < 10 are considered acceptable for path analysis (Kline, 1998). Note also that for the inference questions related to the Revolutionary War and for errors on the fluency measure there are large standard deviations relative to the means.

Table 18

Descriptive Statistics on the Measures for the Entire Sample and the Think-Aloud Sub-Sample

Measure	<i>Entire sample</i>				<i>Think-aloud</i>			
	<i>(N = 175)</i>				<i>sub-sample</i>			
					<i>(n = 44)</i>			
	<i>M</i>	<i>SD</i>	<i>Skew.</i>	<i>Kurt.</i>	<i>M</i>	<i>SD</i>	<i>Skew.</i>	<i>Kurt.</i>
COMP (Raw score; Max = 48)	35.11	10.21	-.92	-.11	33.98	10.14	-1.11	.39
COMPLIT (Max = 28)	20.87	6.21	-.94	-.16	20.11	6.47	-.98	.03
COMPINF (Max = 20)	14.24	4.35	-.79	-.08	13.86	3.94	-.98	.36
COMP (percentile)	58.07	30.55	-.39	-1.07	54.45	28.68	-.44	-.89
BKGD (Max = 20)	13.66	4.48	-.65	-.36	12.11	4.91	-.03	-1.20
BKDDRC (Max = 13)	9.32	2.97	-.79	.03	8.43	3.36	-.20	-1.06
BKDDTA (Max = 7)	4.34	1.97	-.31	-1.01	3.68	1.96	.20	-1.05
INF (Max = 16)	9.41	4.02	-.08	-1.24	8.50	3.66	.20	-1.11
INFRC (Max = 10)	5.92	2.46	-.05	-1.08	5.34	2.38	.30	-1.00
INFTA (Max = 6)	3.50	1.91	-.20	-1.17	3.16	1.68	-.14	-.82
STRAT (Max = 16)	7.72	3.23	.18	-.72	6.64	2.99	.12	-.49
STRATRC (Max = 10)	4.95	2.48	.10	-.79	4.25	2.21	.37	-.57
STRATTA (Max = 6)	2.77	1.26	-.11	-.74	2.39	1.24	-.03	-.62
VOCAB (Max = 34)	22.29	6.30	-.37	-.47	21.07	6.31	-.42	-0.08
VOCABG (Max = 23)	16.71	4.69	-.60	-.32	16.07	4.77	-.67	.38
VOCABTA (Max = 11)	5.58	2.06	-.06	-.47	5.00	1.99	-.07	-.43
FLUENCY (Raw wpm)	138.49	23.93	.00	1.75	132.43	25.54	-.41	-.49

Measure	<i>Entire sample</i>				<i>Think-aloud</i>			
	<i>(N = 175)</i>				<i>sub-sample</i>			
					<i>(n = 44)</i>			
	<i>M</i>	<i>SD</i>	<i>Skew.</i>	<i>Kurt.</i>	<i>M</i>	<i>SD</i>	<i>Skew.</i>	<i>Kurt.</i>
ERRORS	4.27	4.57	.50	7.78	4.77	5.01	2.39	6.32
FLUENCY (Adjusted wpm)	134.22	25.08	-.08	1.67	127.66	25.29	-.39	-.61
LWI (Raw score; Max = 57)	49.38	3.89	-.98	1.04	48.86	4.25	-.57	-.12
WA (Raw score; Max = 30)	21.04	4.87	-1.26	1.42	20.59	5.55	-1.07	.88
LWI (W score; Max = 589)	527.29	15.38	-.27	.39	525.59	16.94	.18	.30
WA (W score; Max = 540)	504.76	11.36	-.69	1.12	503.95	13.04	-.60	.47
WORD (W score; Max = 564.5)	516.03	12.48	-.58	.36	514.77	14.03	-.19	.07
WORDPC	170.04	24.98	-.38	1.05	164.19	25.98	-.35	-.79

Note: Variables used in the path model are shown in **bold**. COMP = Gates-MacGinitie reading comprehension; COMPLIT = literal questions from the Gates-MacGinitie; COMPINF = inference questions from the Gates-MacGinitie; BKGD = researcher-developed background knowledge measure, BKGDRC = questions from BKGD related to the Gates-MacGinitie; BKGDTA = questions from BKGD related to the think-aloud text (TA); INF = researcher-developed inference measure; INFRC = questions from INF related to Gates-MacGinitie comprehension; INF TA = questions from INF related to the TA text; STRAT = Researcher-developed strategy use measure; VOCAB = vocabulary measure combining Gates-MacGinitie and researcher-developed items; VOCABG = odd items from Gates-MacGinitie vocabulary test included in VOCAB; VOCABTA = researcher-developed items from VOCAB related to TA; FLUENCY = one-minute oral reading measure; LWI = Letter-Word Identification subtest of the Woodcock Diagnostic Reading Battery (WDRB); WA = Word Attack subtest of the WDRB; WORD = mean of W scores on LWI and WA; WORDPC = Principal Components composite score calculated from LWI, WA, FLUENCY (adjusted wpm), and ERRORS.

Bivariate scatterplots of all possible pairs of variables were generated as a way to identify outliers; none were immediately evident. Intraclass correlations were calculated using Maximum Likelihood estimation with the SPSS VARCOMP procedure; results are shown in Table 19. ICCs were large and variable for both class and teacher effects; responses on all test measures were substantially similar within classes, potentially biasing the significance tests of the path coefficients. The high ICCs are not surprising, given the academic tracking policies used in the county from at least middle school onwards. One approach to accommodating this non-independence in the data is to use a more conservative alpha in testing the significance of the path coefficients.

Table 19

Sources of Variance and Intraclass Correlations for Class and Teacher Effects

	BKGDRC	INFRC	STRATRC	VOCRC	WORDPC	COMP
Class effects						
Source						
Class	2.49	1.92	1.29	6.27	135.45	29.58
Error	6.18	4.00	4.73	15.38	472.38	73.57
ICC	29%	32%	21%	29%	22%	29%
Teacher effects						
Source						
Teacher	.02	<.00	.11	3.44	.00*	49.23
Error	8.59	5.80	6.00	21.02	615.26	94.83
ICC	0%	0%	2%	14%	0%	34%

* Estimate set to zero by SPSS

Differences between the think-aloud sub-sample and participants who were not selected for the think-alouds were tested with independent sample t-tests (see Table 20). Most of the variables were non-significantly different between the 44 participants in the think-aloud sub-sample and the 131 participants who were not selected, however, significantly lower scores were found for the think-aloud sub-sample on vocabulary related to the think-aloud text and adjusted words per minute (approximately a 5% difference) and on background knowledge and strategies (approximately a 15% difference). Because the correlational analyses of the think-alouds are based on the entire range of scores, these differences in mean scores are not of great concern (see p. 162), as they would be if the analyses used e.g., t-tests of mean differences.

Table 20

Descriptive Statistics and Results of T-tests Comparing the Think-Aloud Sample to Participants who Were not Selected for Think-Alouds

Measure	<i>Think-aloud sub-sample</i> (<i>n</i> = 44)		<i>Remainder of the sample</i> (<i>n</i> = 131)		<i>t</i>	<i>P</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
COMP	33.98	10.14	35.48	10.24	.85	.40
COMPLIT	20.11	6.47	21.12	6.12	.93	.35
COMPINF	13.86	3.94	14.36	4.48	.66	.51
COMP (percentile)	54.45	28.68	59.26	31.16	.90	.37
BKGD	12.11	4.91	14.17	4.23	2.68	.01*
BKDDRC	8.43	3.36	9.61	2.78	2.31	.02*
BKDDTA	3.68	1.96	4.56	1.94	2.59	.01*

Measure	<i>Think-aloud sub-</i>		<i>Remainder of</i>		<i>t</i>	<i>P</i>
	<i>sample</i>		<i>the sample</i>			
	<i>(n = 44)</i>		<i>(n = 131)</i>			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
INF	8.50	3.66	9.71	4.10	1.75	.08
INFRC	5.34	2.38	6.11	2.46	1.80	.07
INFTA	3.16	1.68	3.61	1.98	1.47	.15§
STRAT	6.64	2.99	8.08	3.23	2.62	.01*
STRATRC	4.25	2.21	5.19	2.53	2.20	.03*
STRATTA	2.39	1.24	2.89	1.24	2.36	.02*
VOCAB	21.07	6.31	22.69	6.26	1.49	.14
VOCABG	16.07	4.77	16.92	4.66	1.04	.30
VOCABTA	5.00	1.99	5.77	2.05	2.19	.03*
FLUENCY (Raw wpm)	132.43	25.54	140.49	23.13	1.95	.05
ERRORS	4.77	5.01	4.10	4.42	-.85	.40
FLUENCY (Adjusted wpm)	127.66	25.29	136.39	24.72	2.02	.04*
LWI (Raw score)	48.86	4.25	49.55	3.77	1.01	.31
WA (Raw score)	20.59	5.55	21.19	4.63	.71	.48
LWI (W score)	525.59	16.94	527.86	14.84	.85	.40
WA (W score)	503.95	13.04	505.03	10.78	.54	.59
WORD (W score)	514.77	14.03	516.45	11.94	.77	.44
WORDPC	164.19	25.98	172.00	24.43	1.80	.07

* $p < .05$

§ degrees of freedom adjusted due to unequal variance between groups.

Note: Variables used in the path model are shown in **bold**. See Table 18 for abbreviations.

Correlations and reliabilities in the observed data are shown in Table 21;

Cronbach's alpha reliabilities of all variables entered in the model were adequate ($> .7$).

Reliability of the think-aloud-related strategy questions was low (.28).

Table 21

Correlations and Reliabilities for Reading Measures

	1	2	3	4	5	6	7	8
1. BKGD	.863							
2. BKGDRC	.940	.809						
3. BKGDTA	.858	.630	.729					
4. INF	.724	.667	.643	.830				
5. INFERRC	.662	.617	.574	.938	.704			
6. INFERTA	.672	.608	.613	.896	.686	.744		
7. STRAT	.609	.566	.533	.772	.728	.686	.711	
8. STRATRC	.585	.539	.516	.763	.741	.652	.936	.690
9. STRATTA	.412	.389	.351	.476	.408	.476	.722	.433
10. VOC	.798	.738	.703	.753	.668	.724	.646	.608
11. VOCRC	.786	.725	.696	.727	.650	.691	.643	.601
12. VOCTA	.651	.607	.565	.649	.562	.641	.513	.491
13. LWI	.586	.568	.477	.546	.505	.497	.452	.445
14. WA	.479	.466	.385	.364	.320	.353	.298	.267
15. RAWWPM	.488	.460	.418	.479	.453	.424	.495	.482
16. ERR	-.402	-.378	-.343	-.322	-.263	-.339	-.222	-.237
17. ADJWPM	.539	.508	.461	.516	.480	.466	.512	.503
18. WORDPC	.581	.552	.490	.540	.500	.491	.516	.501
19. COMP	.750	.715	.627	.750	.663	.724	.601	.586
20. CMPLIT	.720	.693	.593	.726	.633	.711	.577	.557
21. CMPINF	.732	.690	.626	.725	.653	.685	.587	.581

Table 21 (cont'd.)

	9	10	11	12	13	14	15	16
1. BKGD								
2. BKGDRC								
3. BKGDTA								
4. INF								
5. INFERRC								
6. INFERTA								
7. STRAT								
8. STRATRC								
9. STRATTA	.277							
10. VOC	.462	.898						
11. VOCRC	.467	.972	.880					
12. VOCTA	.349	.846	.698	.653				
13. LWI	.286	.677	.660	.569	.815			
14. WA	.241	.503	.496	.408	.771	.839		
15. RAWWPM	.320	.533	.504	.483	.578	.517	---	
16. ERR	-.103	-.432	-.438	-.325	-.455	-.412	-.161	---
17. ADJWPM	.324	.587	.561	.521	.635	.569	.984	-.336
18. WORDPC	.338	.637	.612	.555	.767	.722	.951	-.374
19. COMP	.388	.774	.764	.629	.563	.434	.534	-.424
20. CMPLIT	.385	.736	.725	.603	.540	.416	.532	-.409
21. CMPINF	.362	.766	.759	.615	.552	.425	.492	-.413

Table 21 (cont'd.)

	17	18	19	20	21
1. BKGD					
2. BKGDRC					
3. BKGDTA					
4. INF					
5. INFERRC					
6. INFERTA					
7. STRAT					
8. STRATRC					
9. STRATTA					
10. VOC					
11. VOCRC					
12. VOCTA					
13. LWI					
14. WA					
15. RAWWPM					
16. ERR					
17. ADJWPM	---				
18. WORDPC	.976	---			
19. COMP	.587	.620	.943		
20. CMPLIT	.583	.600	.977	.912	
21. CMPINF	.545	.597	.953	.953	.862

Note: Cronbach's alpha reliabilities are shown in *italics* on the diagonal and correlations are shown below the diagonal. Variables used in the path model are shown in **bold**. All correlations above .17 are significant at $p < .01$; all correlations above .13 are significant at $p < .05$. Dashes indicate that reliability could not be calculated for the fluency measures. See Table 18 for abbreviations.

Evidence for validity of the measures. Five types of evidence are available for this study to support the validity of the measures (Crocker & Algina, 1986):

1. *Content validity* based on the domain specifications for the measures (i.e., they were uniformly based on the content of the Gates-MacGinitie comprehension sub-test and the think-aloud passage) and the committee's comments on the measures.

2. *Concurrent validity* based on the large, significant correlations among the measures which are expected (based on the literature reviewed above) and the similar patterns of correlations found in the pilot study and the dissertation study. In addition, none of the measures suffered from restriction of range, which would distort these correlations.

3. *The reliability of the measures* in the pilot sample and the dissertation sample, which is a necessary, but not sufficient condition for validity.

4. *Predictive validity* in that students in honors classes had significantly higher mean scores on every measure than did students in non-honors classes.

5. *Convergent validity* with the think-aloud data, which is presented below (see p. 189).

Research Question 1: Using a new sample, which has the best fit to the data: the CI, VE, or IM model?

I used Maximum Likelihood estimation in EQS 6.1 (Bentler, 1995) to conduct a path analysis on the effect of the 5 reading components (background knowledge, inference, strategies, vocabulary, and the word reading composite) on reading comprehension, testing the three different theoretical models presented above. The path analysis was conducted on the raw data set. Fit indices for the four models are shown in

Table 22. The IM model again had the best fit of the three models, using this new sample and slightly different strategy and word reading measures.

Table 22

Fit Indices for the CI, VE, and IM Models

	<i>CI</i>	<i>VE</i>	<i>DIME</i>
χ^2 (df)	183.132* (7)	175.083* (7)	34.493* (3)
AIC	169.132	161.083	28.493
CFI	.730	.742	.952
SRMR	.196	.190	.085
R ²	.399	.421	.657

* $p < .05$

Note: Figures in **bold** meet Hu and Bentler's (1999) criteria for good fit.

Research Question 2: What is the best-fitting of four related IM models for 9th grade readers?

I again used Maximum Likelihood estimation in EQS6.1 (Bentler, 1995) to conduct a path analysis on the effect of the 5 reading components (background knowledge, inference, strategies, vocabulary, and the word reading composite) on reading comprehension, testing the four variations on the IM model presented above. The path analysis was conducted on the raw data set. Fit indices for the four models are shown in Table 23.

Table 23

Fit Indices for the Four Variations on the IM Model

	<i>Model 1</i>	<i>Model 2</i>	<i>Model 3</i>	<i>Model 4</i>
χ^2 (df)	104.470* (4)	34.493* (3)	140.669* (4)	93.187* (3)
AIC	96.470	28.493	132.669	87.187
CFI	.846	.952	.790	.862
SRMR	.178	.085	.293	.268
RMSEA (CI)	.382 (.319, .445)	.247 (.177, .323)	.446 (.382, .508)	.418 (.346, .490)
R ²	.616	.657	.570	.613

* $p < .001$

Note: Figures in **bold** meet Hu and Bentler's (1999) criteria for good fit.

Since none of the models are nested in each other, they may only be compared non-statistically, using the AIC.¹ The AIC depends on the degrees of freedom in the model and the chi-square statistic, which in turn depends on the sample size and the difference between the observed and model-implied (reproduced) variances and covariances (i.e., the residual variance/covariance matrix). With small differences in chi-square values between models, the number of degrees of freedom in the model can affect the ranking of the models. However, in this case, the chi-square values are so different between models, and the degrees of freedom are so close, that only the residual variance/covariance matrix affects the ranking of the results. The use of non-statistical

¹ Research is underway to develop statistical tests for comparing non-nested models. These approaches use, for example, Markov Chain Monte Carlo methods and/or Bayesian approaches (see e.g., Huelsenbeck, Larget, & Alfaro, 2004).

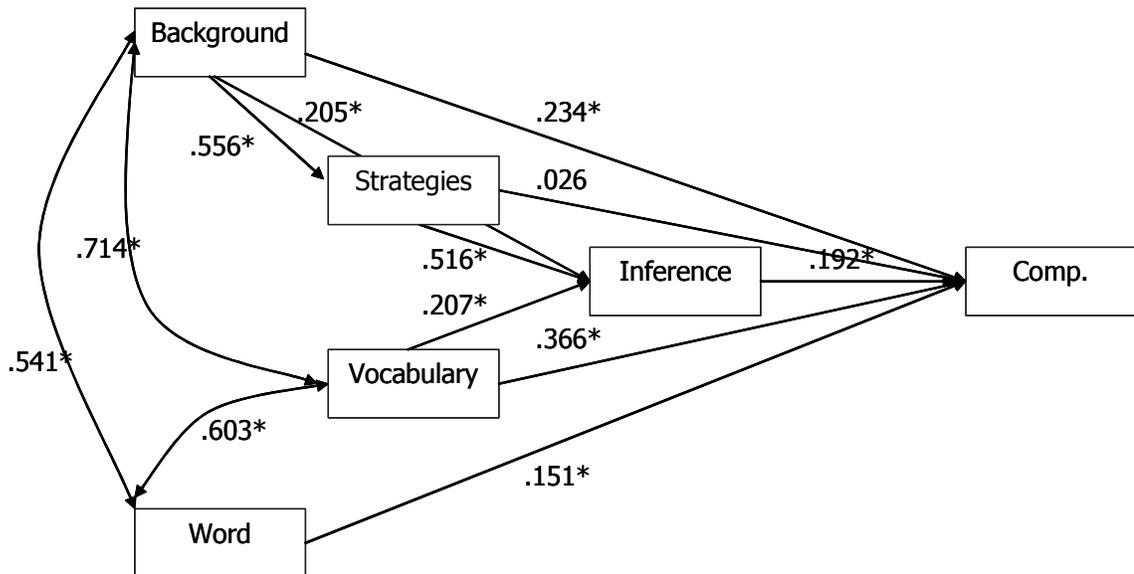
comparisons does technically limit the generalizability of the results from this sample to the population.

Model 2 (the original IM Model) has the best (smallest) AIC of the four variations, followed by Model 4, Model 1, and Model 3. In addition, Model 2 had the best fit by all 6 criteria, it is just at Hu and Bentler's (1999) cutoffs for acceptable fit (CFI $\geq .95$ and SRMR $\leq .09$ or $.10$; it is the only one of the four models to meet those criteria), and it accounts for 66% of the variance in comprehension scores. Model 2 does have a poor RMSEA, probably due to the small sample size.

The best-fitting model, Model 2, with final standardized path loadings, is shown in Figure 8. All paths in the model are significantly different from zero by a one-tailed z test, except for the direct effect of strategies on comprehension.

Figure 8

Final Standardized Solution for Model 2



* Indicates path was significant with a one-tailed z test at $p < .05$; correlations were tested with two-tailed tests.

Research Question 3: What are the predictor variables that make the largest total contribution to reading comprehension in the best-fitting model for 9th grade readers?

The direct and indirect paths showing the effect of each variable on comprehension for Model 2 are shown in Table 24.

Table 24

Standardized Direct and Indirect Paths from Predictors to Reading Comprehension in Model 2

Variable	<i>Direct</i>	<i>p*</i>	<i>Indirect</i>	<i>p*</i>	<i>Total</i>	<i>p*</i>
BKGD	.234	<.00	.109	<.00	.342	<.00
INF	.192	<.00	—	NA	.192	<.00
STRAT	.026	.35	.099	<.01	.125	.01
VOC	.366	<.00	.040	.02	.406	<.00
WORD	.151	<.00	—	NA	.151	<.00

* one-tailed z test.

Note: A dash indicates that the path is not included in the model.

In this model, vocabulary and background knowledge made the largest total contributions to comprehension. Vocabulary had a direct effect that is considered large (using Cohen's [1983] criteria) and a small, but significant, indirect effect via its effect on inference, for a large total effect. This is consistent with the findings from the preliminary study (Cromley & Azevedo, 2004a), in which vocabulary also made the largest total contribution to comprehension. Background knowledge had a medium direct effect on comprehension and a small indirect effect via its effect on inference, for a medium-to-

large total effect. Inference, word reading, and strategies each made a smaller total contribution to comprehension. Inference and the word reading composite had medium-sized direct effects. Strategies had a small and non-significant direct effect and a small and significant indirect effect via their effect on inference.

Research Question 4: How do high- and low-comprehending readers differ on those predictor variables?

Participants were divided into high- and low-comprehending groups based on a median split on Gates-MacGinitie comprehension scores (split at a raw comprehension score of 38; $n = 83$ high comprehenders and $n = 92$ low comprehenders; all comparisons were also significant for analyses where the sample was split at the 50th percentile on the Gates-MacGinitie). A series of independent sample t-tests on the five predictor variables was used to investigate differences across comprehension groups. Group means and results of the t-tests are shown in Table 25.

Table 25

Group Means, Standard Deviations, Results of Independent Sample T-tests, and Effect Sizes for Predictors and for Comprehension Across Low- and High-Comprehending 9th Grade Students

Variable	<i>Low-</i>	<i>High-</i>	<i>t (df)^a</i>	<i>ES</i>
	<i>Comprehending Students (n = 92)</i>	<i>Comprehending Students (n = 83)</i>		
COMP	27.62 (8.74)	43.20 (2.77)	16.22 (110.87)	3.09
COMPLIT	16.54 (5.66)	25.57 (1.81)	14.48 (111.17)	2.75
COMPINF	11.08 (3.61)	17.64 (1.66)	15.69 (130.91)	2.77
COMPCTIL	33.54 (20.70)	84.57 (10.95)	20.65 (141.15)	3.50
BKGD	10.80 (3.92)	16.71 (2.64)	11.78 (160.533)	1.88
BKDDRC	7.59 (2.81)	11.16 (1.76)	10.17 (154.59)	1.66
BKDDTA	3.22 (1.81)	5.55 (1.34)	9.78 (166.73)	1.53
INF	6.66 (2.83)	12.36 (2.83)	13.30 (173.00)	2.04
INFRC	4.36 (1.81)	7.59 (1.90)	11.51 (173.00)	1.76
INFTA	2.30 (1.55)	4.77 (1.36)	11.14 (173.00)	1.71
STRAT	5.70 (2.23)	9.90 (2.69)	11.31 (173.00)	1.74
STRATRC	3.41 (1.69)	6.60 (2.11)	10.97 (157.20)	1.76
STRATTA	2.28 (1.08)	3.30 (1.23)	5.83 (173.00)	.88
VOCAB	18.05 (4.93)	26.89 (4.03)	12.90 (173.00)	2.00
VOCABG	13.55 (3.90)	20.13 (2.65)	13.16 (161.14)	2.09

Variable	<i>Low-</i>	<i>High-</i>	<i>t (df)^a</i>	<i>ES</i>
	<i>Comprehending</i>	<i>Comprehending</i>		
	<i>Students</i>	<i>Students</i>		
	<i>(n = 92)</i>	<i>(n = 83)</i>		
VOCABTA	4.50 (1.67)	6.76 (1.81)	8.61 (173.00)	1.31
FLUENCY	127.73 (22.44)	150.08 (19.99)	6.93 (173.00)	1.06
ERRORS	5.48 (5.25)	2.96 (3.29)	3.83 (154.85)	.63
FLUADJ	122.25 (23.20)	147.12 (20.43)	7.49 (173.00)	1.15
LWIRAW	47.54 (3.96)	51.41 (2.61)	7.69 (158.78)	1.22
WARAW	19.42 (5.67)	22.83 (2.92)	5.07 (139.07)	.86
LWIW	519.96 (14.24)	535.42 (12.22)	7.67 (173.00)	1.17
WAW	501.25 (13.11)	508.65 (7.36)	4.66 (145.93)	.77
BASIC	510.60 (12.81)	522.04 (8.86)	6.92 (162.41)	1.09
WORDPC	157.79 (23.68)	183.61 (18.65)	8.05 (169.99)	1.23

All t-tests were significant at $p < .001$

^adf other than 173 indicates an adjustment for a significant Levene's test, i.e., non-homogeneous variance in the two sub-samples; there was more variance in the scores on all variables for low comprehenders than for high comprehenders.

Students below the median on comprehension as a group scored significantly lower on all predictor measures: background knowledge, inferencing, strategies, vocabulary, and the word reading composite. They also scored significantly lower on all measures related to the think-aloud text, and on the complete measures. Students who had low comprehension scores tended to also have low scores on all of the component measures, and students who had high comprehension scores tended to also have high

scores on all of the component measures. The effect size estimates are interpreted as showing a large difference on all variables (Cohen, 1983).

Research Question 5: How are those predictors revealed in the think-aloud protocols of 9th grade readers?

In this section, I report first on the raw frequency and proportion of use of each variable coded in the think-alouds, and then report the results of the Spearman rank correlation analyses.

Descriptive results of the think-aloud data. The raw counts for each code by group, as well as proportions of use for each code (the frequency of use for each code for each participant divided by the total number of verbalizations for that participant) are shown in Table 26. Think-aloud participants were divided into high- and low-comprehending groups based on a median split on their Gates-MacGinitie comprehension scores (split at a raw comprehension score of 37; $n = 22$ high comprehenders [mean raw comprehension score = 41.55, $SD = 2.65$] and $n = 22$ low comprehenders [mean raw comprehension score = 26.41, $SD = 9.13$]).

There was great variability in how frequently participants verbalized the different variables: strategies, word reading, and inference together made up more than 90% of verbalizations, while vocabulary and background knowledge together made up less than 10% of verbalizations. Recall, however, that activation of background knowledge was also required in order to make certain inferences (i.e., evaluation, knowledge elaboration, and links); together these five codes (EVAL+, KE+, KE-, LINK+, and LINK-) make up 11%-26% of verbalizations for low- and high-comprehending students, respectively. Recall also that accurate word reading was not coded, and that only explicitly verbalized accurate use of vocabulary was coded. There is therefore a great deal of (accurate)

cognitive activity which cannot be captured by the coding scheme. Note also that rarely-verbalized codes can still have a large effect on comprehension (consider, for example, that just a few instances of *knowledge elaboration* can mean the difference between a factual reading of a text and a well-integrated sophisticated reading of a text), while frequently-verbalized codes (e.g., repeatedly expressing *feeling of knowing*) could have little effect. The proportion of verbalization therefore has an indeterminate relationship with the proportion of variance in comprehension explained by the respective variables.

For low-comprehending students, word reading errors was the variable used the highest proportion of the time (39.3% of verbalizations), followed by accurate uses of strategies (19.1% of verbalizations), accurate inferences (16.7% of verbalizations), and inaccurate uses of strategies (13.3% of verbalizations; see Table 26). For high-comprehending students, accurate inferences was the variable used at the highest proportion (32.3% of verbalizations), followed by accurate uses of strategies (24.5% of verbalizations), word reading errors (23.5% of verbalizations), and inaccurate uses of strategies (7.6% of verbalizations).

Table 26

Total Raw Frequency and Mean Proportion of Use for Each Code in the Think-Aloud Coding Scheme, Across Groups

<i>Code</i>	<i>Low-Comprehending</i>		<i>High-Comprehending</i>	
	<i>Students (n =22)</i>		<i>Students (n =22)</i>	
	<i>Total</i>	<i>Mean</i>	<i>Total</i>	<i>Mean</i>
<i>Code</i>	<i>Frequency</i>	<i>Proportion</i>	<i>Frequency</i>	<i>Proportion</i>
Background knowledge	31	2.4%	30	4.4%
Accurate	22	1.7%	23	3.5%
PKA+	22	1.7%	23	3.5%
Inaccurate	9	.7%	7	.9%
PKA-	8	.6%	7	.9%
ANACH-	1	.1%	0	0%
Inferences	276	19.3%	313	35.8%
Accurate	236	16.6%	280	32.3%
BTF+	5	.3%	2	.2%
EVAL+	124	9.2%	134	16.2%
HYP+	9	.6%	23	2.0%
INF+	66	4.4%	52	5.0%
KE+	22	1.5%	45	6.0%
LINK+	10	.6%	24	2.9%

Note: Codes are defined on pp. 157-161 and also in Appendix O.

Sums for accurate and inaccurate codes are shown in **Bold** type; those for major coding categories are shown in *bold italic*.

<i>Code</i>	<i>Low-Comprehending</i>		<i>High-Comprehending</i>	
	<i>Students (n =22)</i>		<i>Students (n =22)</i>	
	<i>Total</i>	<i>Mean</i>	<i>Total</i>	<i>Mean</i>
	<i>Frequency</i>	<i>Proportion</i>	<i>Frequency</i>	<i>Proportion</i>
Inaccurate	40	2.7%	33	3.5%
INF-	39	2.6%	23	2.5%
KE-	0	0%	5	.5%
LINK-	1	.1%	5	.5%
Strategies	495	32.4%	280	32.1%
Accurate	284	19.1%	185	24.5%
COIS+	5	.4%	4	.5%
FOK+	15	1.5%	16	1.4%
IMAGE+	0	0%	1	.2%
RR+	89	5.8%	55	4.1%
SQ+	16	1.0%	19	2.2%
SUM+	157	10.3%	32	11.1%
TN+	2	.2%	58	5.0%
Inaccurate	211	13.3%	95	7.6%
JOL-	29	2.4%	22	1.9%
NOTHINK-	47	3.5%	3	.3%
SUM-	135	7.4%	70	5.5%
Vocabulary	94	6.3%	72	6.5%
Accurate				
VOC+	47	3.2%	54	5.0%
Inaccurate				
VOC-	47	3.1%	18	1.5%

<i>Code</i>	<i>Low-Comprehending</i>		<i>High-Comprehending</i>	
	<i>Students (n =22)</i>		<i>Students (n =22)</i>	
	<i>Total</i>	<i>Mean</i>	<i>Total</i>	<i>Mean</i>
	<i>Frequency</i>	<i>Proportion</i>	<i>Frequency</i>	<i>Proportion</i>
Word reading	588	39.3%	233	23.5%
Inaccurate				
OMIT-	79	5.6%	40	4.2%
SC-	125	8.3%	63	6.0%
WORD-	384	25.5%	130	13.3%
Grand total	1,484		1,028	

Compared to low comprehenders, high comprehenders verbalized proportionately almost twice as many accurate inferences (a mean of 32.3% of verbalizations for high comprehenders vs. 16.6% for low comprehenders) and more than one-and-a-half times as often accurately paraphrased vocabulary (5.0% vs. 3.2%). Some accurate inferences required the activation of background knowledge (i.e., evaluation, knowledge elaboration, and links); high comprehenders verbalized these more than twice as often as did low comprehenders (25.1% vs. 11.3%).

Low comprehenders verbalized nearly twice as many word reading errors (25.5% vs. 13.3%) and inaccurately enacted strategies almost twice as often as did high comprehenders (13.3% vs. 7.6%). Low comprehenders more often stated that they did not understand the text (2.4% vs. 1.9%), and much more often stated that they were “not thinking anything” (3.5% vs. 0.3%). Low comprehenders also re-read text more often than did high comprehenders (5.8% vs. 4.1%) and less often took notes (.2% vs. 5.0%).

Overall, a higher proportion of high comprehenders' verbalizations were accurate codes (65.3% vs. 40.9%), whereas for low comprehenders a higher proportion were inaccurate codes (59.1% vs. 34.7%). As in the pilot study, high comprehenders were not infallibly accurate, and low comprehenders were not totally inaccurate. Rather, each group showed a mix of accurate and inaccurate uses of strategies, and more or less partial mastery of all of the components.

To give a flavor of the think-aloud protocols, excerpts from coded transcripts from a low-comprehending (28th percentile on the Gates-MacGinitie) and a high-comprehending (63rd percentile) student are shown below. See pp. 157-161 for the codes and Appendix O for definitions and examples.

Example 1

Excerpt from coded transcript from a low-comprehending (28th percentile) student

<i>Transcript (reading from text is <u>underlined</u>; [brackets] indicate omission)</i>	<i>Code</i>
[Parliament Taxes the Colonies]	OMIT-
<u>To en-force</u>	WORD-
<u>the Proclamation of 1763</u> the— to— <u>and</u>	SC-
<u>protect the colonists from the Indians, Britain declared that it needed an</u>	
<u>army of 10,000 soldiers in the— in North America.</u>	SC-
<u>But who should</u> I think that's a lot.	KE+
<u>But who should pay to support them? Taxpayers in Britain were already</u>	
<u>burdened with the debt . . . from the French and Indian War.</u>	
JC: Say what you're thinking.	
Participant: . . . mm I wonder, like, if they mean money or like	
laundries?	JOL-
<u>The new British Prime Minister, George Grenville, looked from- for</u>	SC-
<u>a way to make the colonies pay more for their own defense.</u> Oh, it is money.	FOK+

<i>Transcript (reading from text is <u>underlined</u>; [brackets] indicate omission)</i>	<i>Code</i>
--	-------------

The Sugar Act

Grenville knew that the Parliament had never directly taxed the colonies.

That's . . . I think the Sugar Act was stupid! EVAL+

The Navigation Acts had only regulated trade so that the colonies would do most of their buying and selling with Britain. I think that's selfish. EVAL+

Colonial merchants were to pay customs duties—charges on foreign imports—in order to sell non-British goods. . . .

JC: Say what you're thinking.

Participant: . . . mm . merchants were to pay customs duties—charges for foreign imports—in order to sell non-British goods. . I don't know! . . . However, merchants usually avoided the duties by . . bribing officials or smuggling. . I think that's bad. SC- EVAL+

Example 2

Excerpt from coded transcript from a high-comprehending (63rd percentile) student

<i>Transcript (reading from text is <u>underlined</u>; [brackets] indicate omission)</i>	<i>Code</i>
--	-------------

Parliament Taxes the Colonies

To enforce the Proclamation of 1763 and protect the colonists from the Indians, Britain declared that it needed an army of 10,000 soldiers in North America. But who should pay to support them?

JC: Don't forget to say what you're thinking.

Participant: They needed a lot of soldiers KE+
because they needed a good defense. INF+

Taxpayers in Britain were already burdened with the debt from the French and Indian War. I don't know. The new British Prime Minister, George Grenville, looked for a way to make the colonies pay more for their own defense. So they were trying to get people to give them more money. SUM+

The Sugar Act. Grenville knew that the Par-lie-uh-ment WORD-

<u>had never directly taxed the colonies.</u> I don't know what that is.	VOC-
<u>The Navigation Acts had only regulated trade so that the colonies would do almost a- <u>most</u></u>	SC-
<u>of their buying and selling with Britain.</u> So they wanted to have good trade with them.	SUM+
<u>Colonial merchants were to pay customs duties-charges on foreign imports -in order to sell non-British goods. However, merchants usually avoided the duties by bribing officials or smuggling.</u> It's like illegal acts, right.	VOC+

Descriptive results of the verbal recall data. Verbal recall scores ranged from 0 to 50, with a mean of 14.61 and a standard deviation of 12.22. Scores were skewed due to two participants with scores of 50 and 59 (skewness = 1.83, kurtosis = 4.04). To give a flavor of the verbal recall protocols, excerpts from coded transcripts from a low-scoring (1 point) and a high-scoring (30 points) student are shown below.

Example 1

Excerpt from free recall protocol from a low-scoring (1 point) student

Well, that they buried people alive [ERR] if they didn't pay their taxes [ERR] and they- some of them were hanged [ERR] and that's about it. . . .
uh . they passed the *Sugar Act*.

Note: Major topics (+4 points each) are shown in *italics and underlining*; sub-topics (+2 points each) in *italics*; and supporting details (+1 point each) in underlining. Errors (-1 point each) are indicated by [ERR].

Example 2

Excerpt from free recall protocol from a high-scoring (30 points) student

Britain needed to pay back money from . . .the French and Indian War because it had put them into debt. So they were figuring out ways to tax

the colonies, and one way was, uh, I think it was the Sugar Act, and it, uh, lowered the cost for getting *molasses*, but it *cracked down* on people who were importing it illegally. And, uh, people were *protesting* against that. But then the major one was the Stamp Act, and it put official stamps on things such as playing cards and newspapers, and it, uh, . . . everything [ERR] like it had, *you had to pay a tax on the stamp*, so uh, people were protesting and they- sometimes in violent ways. For example, they tried to bury a guy alive, so awful! And also they boycotted, they didn't buy the stamped products. And instead they made their own or went without [ERR] . . . They were called the Sons and Daughters of Liberty, yes they were. And , oh, in Britain in the Parliament they were having some *disagreements*, some people thought that America should be grateful to Britain because they were like the children of Britain. But then other people were- said that uh, the reason the Americans were here is- one man said the reason the Americans were here is that uh, was because Britain pushed us out [ERR] and we were forced to flee to this cold and inhospitable land and we had made it our own and they didn't have any right to be- to be taxing us.

Think-Aloud Results as Convergent Validity Evidence for the Paper-and-Pencil Measures. If the paper-and-pencil and think-aloud Revolutionary War measures are in fact tapping the same underlying knowledge and skills, then we should expect high correlations between them. We should expect high, significant, positive Spearman rank correlations between each paper-and-pencil Revolutionary War measure and its corresponding think-aloud code for accurately enacting the variable. Likewise, we should

expect high, significant, negative correlations between each paper-and-pencil Revolutionary War measure and its corresponding think-aloud code for inaccurately enacting the variable. Table 27 summarizes this validity evidence in Spearman rank correlations between scores on the think-aloud related paper-and-pencil measures and verbalizations from the 44 participants who provided think-alouds.

The results provide partial evidence of convergent validity: for four of the five predictor measures and for comprehension at least one of the correlations was significant. High scores on the paper-and-pencil Revolutionary War background knowledge measure were significantly and positively correlated with a higher proportion of verbalizing accurate background knowledge when reading the Revolutionary War text. High scores on the paper-and-pencil Revolutionary War inference measure were significantly and positively correlated with a higher proportion of verbalizing accurate inferences when reading the Revolutionary War text. High scores on the paper-and-pencil Revolutionary War vocabulary measure were significantly and negatively correlated with a higher proportion of verbalizing a lack of understanding of vocabulary when reading the Revolutionary War text. This means, for example, that participants who had high scores on the Revolutionary War vocabulary measure rarely verbalized a misunderstanding of a vocabulary word in the think-alouds, whereas those with low scores on the vocabulary

Table 27

Spearman Rank Correlations Between Scores on Each Component Measure and the Proportion of Verbalizations in the Corresponding Think-Aloud Variable

<i>Scores on Component Measure</i>	<i>Spearman Rank Correlation with Proportion of Corresponding Think-Aloud Code</i>	
	<i>Accurate</i>	<i>Inaccurate</i>
	<i>Verbalizations</i>	<i>Verbalizations</i>
<i>Revolutionary War Specific</i>		
Background Knowledge	.40*	.11
Inference	.38*	-.06
Strategies	.03	-.09
Vocabulary	-.03	-.39*
<i>General</i>		
Word reading composite	NA	-.64*
Comprehension	.57*	NA

* Significant at $p < .05$ by a two-tailed test.

measure more frequently verbalized misunderstandings of a vocabulary word in the think-alouds. High scores on the word reading composite were significantly and negatively correlated with a higher proportion of word reading miscues when reading the Revolutionary War text. Recall that reliability for strategy questions related to the Revolutionary War was unacceptably low, so the non-significant correlation with strategies verbalized during the think-aloud should be interpreted with caution.

In addition, free recall scores should have significant Spearman rank correlations with comprehension scores (as well as the various think-aloud codes; see below). Gates-MacGinitie comprehension and free recall scores had a significant $r_s(43) = .40, p < .05$.

Results of the Spearman rank correlation analyses on the think-aloud data.

Spearman rank correlations were performed on the proportion of verbalization of each accurate and inaccurate code across participants for the five predictor variables. These correlations are then compared to the correlations underlying the path analysis. The analysis therefore does not provide evidence of an effect, it simply suggests that parallel relationships may exist in the paper-and-pencil and think-aloud data.

Correlations among accurate verbalizations (e.g., between paraphrasing vocabulary definitions and making accurate inferences) and with free recall are shown in Table 28.

Table 28

Spearman Rank Correlations Among Proportions of Accurate Verbalization of the Predictor Variables and Verbal Recall

Code	BKGD+	INF+	STRAT+	VOC+
BKGD+	—	$N = 44$		
INF+	.54*	—		
STRAT+	-.22	-.36*	—	
VOC+	-.04	-.12	NA	—
VERBREC	.30*	.39*	.07	.25

* $p < .05$

NA = A relationship not hypothesized in the model

Among codes for verbalization of accurate use of the variables, four of the nine relevant correlations were significant. Activation of accurate background knowledge was significantly and positively associated with accurate inferences. Accurate use of strategies was also significantly but negatively associated with accurate inferences (contrary to expectations). Activation of accurate background knowledge and accurate inferences were significantly and positively correlated with scores on the free recall protocol.

Correlations among inaccurate verbalizations (e.g., between activating inaccurate background knowledge and inaccurately enacting a strategy such as summarizing) and with free recall are shown in Table 29.

Table 29

Spearman Rank Correlations Among Proportions of Inaccurate Verbalization of the Predictor Variables and Verbal Recall

Code	BKGD-	INF-	STRAT-	VOC-	WORD-
BKGD-	—				
INF-	.47*	—			
STRAT-	.08	-.20	—		
VOC-	.08	.03	NA	—	
WORD-	-.07	NA	NA	.35*	—
VERBREC	-.02	.16	-.17	-.33*	-.43*

* $p < .05$

NA = A relationship not hypothesized in the model

Among codes for verbalization of inaccurate use of the variables, four of the twelve relevant correlations were significant. Activation of inaccurate background knowledge was significantly and positively associated with inaccurate inferences. Word reading inaccuracy was significantly and positively associated with misunderstanding vocabulary. Both word reading inaccuracy and misunderstanding vocabulary were significantly and negatively correlated with scores on the free recall protocol.

Correlations between accurate and inaccurate verbalizations (e.g., mis-reading a word, but then making an accurate inference anyway) are shown in Table 30.

Table 30

Spearman Rank Correlations Between Proportions of Accurate and Inaccurate Verbalization of the Five Predictor Variables

Code	BKGD+	INF+	STRAT+	VOC+
BKGD-	NA	-.03	-.04	.16
INF-	.40*	NA	-.07	-.02
STRAT-	-.32*	-.52*	NA	NA
VOC-	-.34*	-.34*	NA	NA
WORD-	-.32*	NA	NA	-.43*

* $p < .05$

NA = A relationship not hypothesized in the model

Between codes for verbalization of accurate and inaccurate use of the variables, seven of 12 relevant correlations were significant. All of these correlations are expected to be negative, since accurate use of one variable should be negatively associated with inaccurate use of another variable. Inaccurate use of strategies was significantly

negatively associated with activation of accurate background knowledge and accurate inferences. Misunderstanding of vocabulary was also significantly negatively associated with activation of accurate background knowledge and accurate inferences. Word reading errors were significantly negatively associated with accurate use of activation of accurate background knowledge and understanding vocabulary. Activation of accurate background knowledge was significantly and positively associated with inaccurate inferences (contrary to expectations).

Table 31

Summary of Significant Correlations among the Predictor Variables from the Paper-and-Pencil and Think Aloud Data

Code	BKGD	INF	STRAT	VOC	WORD
INF					
P&PRC	.62				
P&PTA	.61				
TA	.54++ .47-- .40+ -				
STRAT					
P&PRC	.54	.74			
P&PTA	.35	.48			
TA	-.32+ -	-.36++ -.52 + -			
VOC					
P&PRC	.72	.65	NH		
P&PTA	.56	.64	NH		
TA	-.34+ -	-.34 + -	NH		
WORD					
P&PRC	.55	NH	NH	.61	
P&PTA	—	NH	NH	—	
TA	-.32+ -	NH	NH	.35-- -.43+ -	
COMP					
P&PRC	.72	.66	.59	.76	.61
P&PTA	—	—	—	—	—
TA	.30++	.39++		-.33--	-.43--

Note: BKGD = Background knowledge, INF = Inference, STRAT = Strategies, VOC = Vocabulary, Word = Word reading, COMP = Comprehension, P&PRC = Paper-and-pencil subtest relevant to the content of the Gates-MacGinitie, P&PTA = Paper-and-pencil subtest relevant to the content of the think-aloud text, TA = coded think-aloud verbalizations, NH = a relationship not hypothesized in the model. ++ indicates a correlation between two codes for accurate verbalization, -- indicates a correlation between two codes for inaccurate verbalization, +- indicates a correlation between a code for accurate verbalization (shown in the top row of the table) and a code for inaccurate verbalization (shown in the left hand column), -+ indicates a correlation between a code for inaccurate verbalization and a code for inaccurate verbalization. — indicates no corresponding paper-and-pencil measure that is specific to the think-aloud text.

In summary, eleven of the twelve relationships in the think-aloud data showed significant results using Spearman rank correlations. There was contradictory evidence regarding, the relationship between strategies and inferences and the relationship between background knowledge and inference. There was no evidence from the analysis of the think-alouds to support the relationship between strategies and comprehension. Table 31 summarizes the correlations used in the path analysis, correlations among the paper-and-pencil questions, and the significant Spearman rank correlations in the think-aloud data.

Illustrative examples from the think-alouds corresponding to paths in the model.

Below, I present illustrative examples from the think-aloud protocols to illustrate a number of paths in IM Model 2. All student names used below are fictitious.

Effect of background knowledge on strategies. Damian activated prior knowledge, followed by an accurate summary “they decided to boycott—refuse to buy—British goods. And I remember this a long time ago [PKA+], how they wouldn’t buy goods from Britain [SUM+].”

Effect of strategies on inference. Betty followed an accurate summary with an accurate inference: “However, merchants usually avoided the duties by bribing officials or smuggling. Yeah, so they do bribe and smuggle [SUM+]. One product often smuggled from the French West Indians was molasses, which was used for making rum, especially in New England. So then the people who are drunk aren’t the people who are smuggling [KE+].” John followed an inaccurate summary with an inaccurate inference: “Greenville decided that enforcing the duty on foreign molasses would be a good way to raise revenue—income. In 1764 Parliament passed the Sugar Act, which cut in half the duty on foreign molasses to encourage merchants to pay it. JC: Say what you’re thinking.

John: . . . I'm thinking about the revenue, the raise on molasses [SUM-]. Uh, because others were smuggling it, so they put a raise on foreign molasses [INF-].”

Effect of word reading on comprehension. Mary omitted an important paragraph heading, and then expressed a *judgment of learning*: “[Point of View: Did the Colonies Owe Obedience to Britain?] Even within Parliament there were different points of view on the question. I'm thinking whose points of view they were and, uh, . what questions were raised and all that.” On the other hand, Betsy misread “and” for “by,” but nonetheless expressed a *feeling of knowing*: “New England merchants, who were most affected and the Sugar Act. I know that.”

Effect of vocabulary on inference. Nancy expressed an understanding of vocabulary followed by a correct inference, “Taxpayers in Britain were already burdened with the debit . . . from the French and Indian War. So that they're not, I guess they don't have enough money to [VOC+] . from the previous war that they had [INF+].” On the other hand, Todd expressed an understanding of vocabulary, but followed that with an incorrect inference, “Taxpayers in Britain were already burdened with the debt from the French and Indian War. They owed money [VOC+] because they lost [INF-].”

Effect of word reading on vocabulary. Ki misread a word, followed by a statement of failing to understand the vocabulary word, “this was Pair-lie-ment's [WORD-] first attempt to force the colonies to pay any tax other than customs duties. JC: say what you're thinking. Ki: What does Pair-lie-ments mean? [VOC-]”

In addition to these paths in the model, there was evidence for two other, previously uninvestigated correlations:

Correlation between strategies and vocabulary. Claude gave a summary that included an understanding of vocabulary, and followed this with a hypothesis: “They

sent a petition to Parliament asking it to repeal—do away with—the Stamp Act and the Sugar Act. So they don't like it and they're asking them [SUM+] to stop it [VOC+] and they might go to war if they don't [HYP+]." On the other hand, Bob expressed a misunderstanding of vocabulary, yet nonetheless gave an accurate summary, followed by an accurate inference: "decided to boycott—refuse to buy—British goods. And then that's more of them joining together and . boycotting and striking [VOC-] all their goods [SUM+] so that they don't- instead of getting more money with the taxes, they're going to lose more money, hopefully [HYP+]."

Correlation between strategies and word reading. Jacob misread "opposition" for "oppression" followed by an inaccurate summary: "Your opposition [WORD-] planted them in America. I guess- . . I'm a little confused by that [JOL-], but I guess he means that it wasn't really them, and that they didn't need their help, and all that [SUM-]." On the other hand, Lisa misread "tyranny," and then gave an inaccurate summary: "They fled from your tire-uh-nee to a then uncultivated and inhospitable country. Uh, they left from the tire-uh-nee to a lower country."

Overall, there was a great deal of converging evidence about these students' reading comprehension and components from the paper-and-pencil measures and the think-aloud and free recall data. These two "lenses" do not show a radically different picture of high school students' comprehension or the processes involved in it. Rather, both sources of data suggest that all of the components are important to comprehension, and that they act and interact in complex ways.

CHAPTER VI: DISCUSSION

The Sample and Think-Aloud Sub-Sample

The full sample was larger and more diverse in reading comprehension skills and race than that of many reading comprehension studies using paper-and-pencil measures. As planned, participants spanned the range of reading skills, from 1st to 99th percentile on the comprehension measure, with a mean of 58th percentile. Scores on most measures were slightly negatively skewed, i.e., high scorers were slightly over-sampled. However, this amount of skewness was not so high that it posed a threat to the analyses.

Intercorrelations among the measures were uniformly large, significant, and in the expected direction, suggesting that a multivariate analysis was appropriate. Consistent with the *a priori* power analysis (and supported by statistically significant results) the study design had sufficient power to detect differences in this sample. There was substantial evidence for the validity of the measures; this included Cronbach's alpha reliabilities in the acceptable to excellent range (.69 to .94), except for strategy questions related to the Revolutionary War.

The think-aloud sub-sample was likewise larger and more diverse in reading comprehension skills and race than that of many reading comprehension studies using think-aloud methodology. The sub-sample was quite similar to the students who were not selected to provide think-alouds in terms of their reading comprehension scores, age, and race. However, the think-aloud sub-sample had more females and had slightly slower

mean reading rates, lower background knowledge, reading comprehension strategies, and vocabulary scores related to the think-aloud passage than those not selected for the think-aloud.

Data Screening

Screening showed that the data met the assumptions of normality, but to some extent violated the assumption of independence of observations. That is, within classes and within teachers, students were more like each other than would be expected if students were randomly assigned to classes. We know, however, that this is not the case; students are assigned to honors, regular, and remedial classes on the basis of academic performance (for honors classes) and reading test scores (for the remedial class). This non-independence tends to falsely decrease the standard error for the measures, increasing the risk of Type I error when testing the significance of each path in the model.

Model Comparisons

Replicating the comparison of the CI, VE, and IM Models. Using a new sample, and slightly different measures of strategies and word reading, the IM model still had the best fit of the three theoretical models, and had a dramatically better fit than either the CI or VE models. In addition, the IM model was the only one to meet Hu and Bentler's (1999) criteria for good fit. In the preliminary study, a reliability adjusted model was fit to the data, so comparisons to the non-adjusted model tested in the current study must be interpreted cautiously.¹ Nonetheless, vocabulary made the largest contribution to comprehension in both the preliminary study (whether using a reliability adjusted or non-adjusted approach) and the present study. Likewise, word reading made the smallest

¹ Note that the large proportion of variance explained by the reliability-adjusted model is because all of the error variance in every measure has been entered into the model.

contribution to comprehension in the preliminary study (using a reliability adjusted approach; word reading ranked 4th using a non-adjusted approach) and the present study.

Path coefficients cannot be compared between the reliability-adjusted and non-adjusted models, but a comparison between the path coefficients for the non-adjusted model in the preliminary study and the present study is a fair one. Considering the magnitude of the path coefficients using Cohen's (1983) guidelines, in the present study the effects of background knowledge and word reading are somewhat larger than in the preliminary study (i.e., from a small effect to a medium effect or from a medium effect to a large effect), the effects of vocabulary and inference are somewhat smaller, the correlations among the exogenous variables are slightly smaller, and the effect of strategies on comprehension is somewhat smaller while their effect on inference is somewhat larger. Changes to the measures may have contributed to these changes in the relative size of the path coefficients. This instability in the size of the path coefficients suggests that a replication with a larger sample and the same measures as the current study would be useful.

Refining the IM model by testing four variations. Results of the path analysis showed that Model 2, the original Inferential Mediation Model, had a better fit than any of the other three alternative models. In addition, the CFI and SRMR for Model 2 were in the range of acceptable fit as recommended by Hu and Bentler (1999) and it accounted for a substantial 66% of the variance in comprehension. The better fit of Model 2 suggests that, with these measures and this multi-cultural 9th grade sample, adding the path from word reading to vocabulary does not improve the fit of the model. It is important to keep in mind that there is a nearly infinite number of alternative models that would fit as well or better than Model 2. This study simply provides evidence that Model

2 has the best fit among the four specific models that were proposed. This is consistent with the excellent fit of the original Inferential Mediation model in the preliminary study (Cromley & Azevedo, 2004a), despite using different strategy and word reading measures, and modifying the vocabulary measure. This study is therefore a robust replication of the preliminary study.

Two specific paths were tested in the model comparisons. With regard to the effect of background knowledge on strategies, both models with this path were better fitting than the corresponding model with a correlation instead of a direct effect (Model 2, AIC = 28.49 and Model 4, AIC = 87.19; Model 1, AIC = 96.47 and Model 3, AIC = 132.70). The path coefficient for this effect was large (.556), although comparisons among paths need to be made cautiously because the different measures had different reliabilities. Prior research on the effect of background knowledge on strategies is contradictory. In some studies, higher background knowledge significantly relates to more effective use of strategies (e.g., Symons & Pressley, 1993), and sometimes does not (e.g., Dreher & Brown, 1993). These contradictory findings might be due to the type of strategy or strategies measured, the age group(s) studied, the measures of background knowledge and/or strategy use, or some combination of these. Analyses of the think-aloud protocols do suggest a relationship between inaccurately activating background knowledge and inaccurate use of strategies, especially inaccurate summaries. Students who activated inaccurate or irrelevant background knowledge more often gave muddled summaries of what they had just read.

With regard to the effect of word reading on vocabulary, both models that included this path (Model 1, AIC = 96.47; Model 3, AIC = 132.70) fit worse than Model 2 (AIC = 28.493). There is meager prior support for this path from students in fourth

grade or older. There is a strong theoretical rationale for decoding being required prior to accessing a word in the mental lexicon (although competing theories have been put forth; see, e.g., Coltheart & Rastle, 1994). Despite the worse fit of the two models which include this direct path, there was some support in the think-alouds for the relationship between word reading and vocabulary. On some occasions, participants failed to accurately decode a word (e.g., “repel” for *repeal*), yet they accurately paraphrased the meaning of the word as written (e.g., “repel means do away with”). On the other hand, participants more often decoded inaccurately and then showed evidence that they did not understand the word (e.g., the participant who read “revenge” for *revenue*, and then proceeded to explain that the passage was about the British getting revenge on the colonists). Students also sometimes decoded accurately, but then showed evidence of either understanding or not understanding the vocabulary words in the passage. With regard to the Spearman rank correlations, participants who more often verbalized inaccurate word reading also significantly more often expressed difficulty with or misunderstandings of vocabulary (e.g., “eh-fee-ghee [effigy], I don’t know what that is”). In summary, evidence from the think-alouds to some extent supports and to some extent contradicts findings from the model fitting.

All direct and indirect path loadings as well as correlations in the final model were significant, except for the direct effect of strategies on comprehension. All of these paths are attenuated to some extent by measurement error in the measures, and by the differing reliabilities of the various measures. The significant path loadings are strongly consistent with the large body of prior research that was used to build the model, with the exception of studies showing a direct effect of strategies on comprehension. Consistent with the results of the preliminary study, this suggests that the variables have their effect

on comprehension both directly and indirectly, via inference. Whereas previous studies have for the most part regressed predictors directly on comprehension (e.g., Saarnio et al., 1990), these results suggest that the indirect effects via inference are an important part of comprehension, as in the CI (Kintsch, 1988) and VE (Perfetti, 1985) models, and should not be neglected.

Correlations among the exogenous variables were mostly in the range to be expected from the literature review. In Model 2, the maximum likelihood estimate of the correlation between background knowledge and vocabulary was .71, and the literature review showed Pearson correlations of .36 to .84. The estimate of the correlation between the word reading composite and vocabulary was .60, and the literature review showed Pearson correlations for fluency/accuracy and vocabulary with absolute values of .40 to .61. The estimate of the correlation between background knowledge and the word reading composite was .54, and the literature review showed Pearson correlations for fluency/accuracy and background knowledge with absolute values of .19 to .27. The high correlation in this sample between background knowledge and the word reading composite could be due to differences in how both the background knowledge and word reading constructs were measured, could be due to sampling error, or could be due to the clustering within classes.

As a whole, Model 2 suggests that vocabulary and background knowledge contribute to reading comprehension both directly—when a literal understanding is needed—and indirectly via inference when the text demands that the reader draw logical conclusions. The model further suggests that reading comprehension strategies (as measured in this study) primarily have their significant impact when inferences are needed; that is, the effect of strategies on comprehension is mediated by inference.

Finally, results of the model fitting suggest that the word reading accuracy and fluency composite has only a direct effect for 9th grade students like these.

As in the preliminary study, these results suggest that direct paths from the variables to comprehension are important. While the indirect paths emphasized so strongly by Kintsch (1988, 1998) and Perfetti (1985) are significant in this model, the effects are small ones.

Relative Contribution of the Components to Reading Comprehension

I now consider each variable in turn, in the relative order of their contribution to comprehension. Direct and indirect effects in the model fitting need to be viewed in the context of the different reliabilities for each variable (ranging from .69 to .93). That is, the relative sizes of the effects as estimated in the model fitting fail to take into account differing reliabilities for the measures. The contributions made by the variables are best seen in groups: background knowledge and vocabulary making a larger contribution, and inference, strategies, and word reading making a smaller contribution. For each variable, I first discuss the findings from the paper-and-pencil measures, then the results from the think-alouds, and then the results of the Spearman rank correlations coordinating the two.

Vocabulary. Consistent with prior research on high school students (e.g., Graves et al., 1980), vocabulary had a medium-to-large total effect on comprehension.

Vocabulary had most of its effect directly on comprehension, but also indirectly via inference. This is the first study with high school students to show evidence for this indirect effect, but evidence exists for 4th-8th grade students and undergraduates, e.g., Walczyk and Taylor (1996).

Some examples of difficult vocabulary from the paper-and-pencil vocabulary measure were “officials” (“employers” or “spokesmen” were chosen by 30% of

participants), “enforce” (“force into” was chosen by 42% of participants), and “shudder” (“a cry of fear” was chosen by 28% of participants).

In the think-aloud text, examples of vocabulary that posed difficulties include “peers” (misunderstood to mean ‘friends,’ instead of ‘social equals’), “customs duties” (which was defined in the text, but several participants verbalized that they did not know what the word meant), and “smuggled” (e.g., “Thinking about stealing” or “I think of drugs”).

Both the direct and indirect paths for vocabulary were reflected in the think-alouds, also consistent with previous think-aloud studies with high school students (e.g., Kletzien, 1992). For example, participants often accurately paraphrased vocabulary definitions, e.g., for debt, “they didn’t have enough money.” On the other hand, one participant, after reading about the Stamp Act Congress sending a petition to Parliament, said “they sent someone to Parliament,” showing a lack of understanding that a petition is a written document. Likewise, several participants referred to Parliament as “he” or “him” suggesting that they believed Parliament as a person (perhaps similar to the newspaper convention of referring to people by their last name, as in ‘Clinton asserted that . . . ’).

Results of the Spearman rank correlation analyses provide correlational evidence to corroborate the correlations that underlie the path model: participants who accurately paraphrased vocabulary at higher proportions tended to make a smaller proportion of meaning-affecting miscues. Participants who expressed misunderstandings of vocabulary at higher proportions tended to activate accurate background knowledge at low proportions, make a smaller proportion of accurate inferences, have a higher proportion of miscues, and have lower free recall scores. Note that non-significant correlations (e.g.

between VOC+ and INF+) could be due to the relatively small sample size or the use of less-powerful Spearman rank correlations (compared to Pearson correlations). Also note that the paths in the model are directional, whereas correlations are not; one consequence is that suppressor variables could cause non-significant relationships to appear significant in a correlation, and *vice versa*.

In addition, there was a suggestion in the think-alouds that participants who expressed more difficulty with vocabulary also had difficulty with accurately summarizing what was read; that is, a connection between vocabulary and strategies which is not represented in the model. This relationship is suggested in the *petition* example above. There is a slight suggestion of this relationship in Freebody and Anderson (1983a), in that 6th grade students wrote significantly better two-to-three sentence summaries of social studies passages that had been rewritten with easier vocabulary than those with the original, more difficult vocabulary. Summaries in this context were used as a measure of comprehension, rather than as a measure of strategy use.

Together, these analyses of the paper-and-pencil measures, analyses of the think-alouds, and analyses that coordinate the product and process measures provide converging evidence that vocabulary plays an important role in 9th grade students' comprehension, and that it may contribute to comprehension difficulties for those students who struggle to understand what they read.

Background knowledge. Consistent with prior research on high school students (e.g., Stevens, 1980), background knowledge had a medium-sized total effect on comprehension. Background knowledge had most of its effect directly on comprehension, but also indirectly via strategies and inference (as in McNamara et al., 1996). These

contradictory findings in prior research suggest that we need to identify circumstances under which prior knowledge does and does not have an effect on strategy use, and for which strategies it has an effect.

Some examples of difficulty with background knowledge from the paper-and-pencil measure were identifying the date of the Declaration of Independence (only 60% of participants chose 1776), identifying the parties in the Revolutionary War (32% chose an answer other than the American colonies and Britain), and identifying the fate of rain water (absorption and runoff; answered incorrectly by 32% of participants).

In the think-aloud text, participants sometimes verbalized background knowledge. For example, students stated that the text “reminds me of what we learned about the Revolutionary War last year.” They also recalled that “license plates in the District of Columbia say that, ‘No taxation without representation.’” Participants’ prior beliefs about taxes as people ‘being cheated out of something’ by the government created obstacles for comprehension. As in VanSledright’s (1995) interviews with 8th grade students who had just finished a unit on colonization, these students seemed to have a “factual stew” of disconnected information they remembered from their American history class the previous year.

Both the direct and indirect paths from background knowledge in the model were reflected in the think-alouds, consistent with previous studies with high school students (e.g., Afflerbach, 1990). As an example of the direct path, Louis read about the colonies doing “most of their buying and selling with Britain. And thinking of all the trade and all the traders and that stuff that I learned before.” In the think-alouds, participants also activated background knowledge and then drew inferences between the text and that background knowledge, sometimes accurately and sometimes inaccurately. For example,

Louis followed the above comment with an inference that “people are having to pay too much for the trading and all that stuff.”

Results of the Spearman rank correlation analyses provide correlational evidence to corroborate the correlations that underlie the path model: participants who activated a high proportion of accurate background knowledge tended to make a higher proportion of accurate inferences, but they also tended to make a higher proportion of inaccurate inferences. They also tended to have higher verbal recall scores, have a smaller proportion of inaccurate use of strategies and misunderstandings of vocabulary, and make a smaller proportion of miscues. Participants who activated a high proportion of inaccurate background knowledge tended to make a high proportion of inaccurate inferences. In addition, in the think-alouds participants who activated inaccurate background knowledge also had difficulty accurately summarizing what was read.

Background knowledge might have a greater opportunity to play a role because I used the Gates-MacGinitie reading comprehension subtest, since this test includes both narrative and expository passages across a range of domains, including biology, ecology, and history. On the other hand, an argument could be made that standardized comprehension tests have removed some items that rely on background knowledge (presumed to be race- or class- biased) in response to accusations of bias over the last 30 years (Murphy, Shannon, Johnston, & Hansen, 1998).

Together, these analyses of the product measures, analyses of the process measure, and analyses that coordinate the product and process measure provide converging evidence that background knowledge plays an important role in 9th grade students’ comprehension, and that it may contribute to comprehension difficulties for those students who struggle to understand what they read.

Inference. Consistent with prior research on high school students (e.g., Hare et al., 1989), inference had a medium-sized direct effect on comprehension. Paper-and-pencil inference questions that posed difficulties for participants included identifying a document that would not require a tax stamp (answered incorrectly by 51% of participants), inferring a character's emotional state (answered incorrectly by 55% of participants), and inferring why a turtle was tired (answered incorrectly by 33% of participants).

In the think-alouds, participants made both correct and incorrect inferences across a wide range of inference types, including moral *evaluations* of what they were reading, *inferences* within text and *knowledge elaboration* between prior knowledge and text, *links* with other historical knowledge, and *hypotheses* about what was to come in the text. Across all participants, accurate inferences were 23.9% of verbalizations and inaccurate inferences were 2.6%. An example of an accurate LINK+ comes from Jane, who read about colonial boycotts, and responded “I just learned that in US History!” in a unit on the U.S. Civil Rights Era. An example of an inaccurate INF- was when Abner read about colonists burning an effigy of a tax collector and concluded that “the colonists was trying to send a message that they will do anything to be heard and seen.”

Results of the Spearman rank correlation analyses provide correlational evidence to corroborate the correlations that underlie the path model: participants who had a high proportion of accurate inferences tended to activate a high proportion of accurate background knowledge and tended to have higher free recall scores. They tended to less often inaccurately enact strategies, and less often misunderstand vocabulary. However, they also tended to less often accurately enact strategies. Participants with a high

proportion of inaccurate inferences tended to more often activate inaccurate background knowledge at high proportions.

Clearly, this American history passage provided many opportunities for students to draw (or fail to draw) inferences. This text feature is common of texts that students face as they advance through the grades, and the increasing need for inferences is one of the features of high school texts that make them difficult (Alexander & Jetton, 2000). Like Neuman's (1990) 5th-grade students, low-comprehending participants in this study made more inference errors than did high comprehenders, including over-relying on inaccurate or irrelevant background knowledge.

Together, these analyses of the product measures, analyses of the process measure, and analyses that coordinate the product and process measure provide converging evidence that inference plays an important role in 9th grade students' comprehension, and that it may contribute to comprehension difficulties for those students who struggle to understand what they read.

Word reading. Consistent with prior research on high school students (e.g., Artelt et al., 2001), word reading had a significant direct effect on comprehension. On the real word and pseudoword reading measures, it was infrequent words that participants had the most difficulty with, and these tended to be polysyllabic words, affixed words, and words of foreign (e.g., French) origin. In no case, however, did the students lack the most basic decoding skills—they did not mis-read CVC or CVCE patterns in monosyllabic words or pseudowords; they correctly used the most common letter-sound correspondences for consonants. Rather, to the extent that there was a weakness in word reading accuracy, it occurred with infrequent words with uncommon orthographic patterns, especially ones that reflected changes in morphology (as in final → finalities).

The case is different, however, for word reading fluency. On average, participants read at 135 correct words per minute, close to the average for 9th grade students. The word composite had a medium direct effect on comprehension, consistent with automaticity theories of reading comprehension (e.g., Stanovich, 1988). This effect is consistent with the hypothesis that slow, inefficient word reading takes up mental energy that is needed for the work of comprehension. To use an extreme example, the think-aloud participant with the slowest word reading speed (73 words per minute; this participant also tested at 36th percentile on word reading accuracy) said, “I don’t think when I read, I just read.”

The think-alouds suggested that participants frequently mis-read or omitted words in ways that were likely to affect meaning (30.9% of verbalizations). For example, participants misread “col-o-nel” for *colonel*, “Par-lie-uh-ment” for *Parliament*, and “tire-uh-nee” for *tyranny*. Note that while the first two are phonetic readings of the spellings, the third may be read by analogy to *tyrant*. This suggests a weakness in the relation between morphology (derived words) and phonology (pronunciation patterns) noted by Nagy et al. (1993) in their study of high school students’ vocabulary knowledge.

For word reading accuracy, the results of the paper-and-pencil measures and of the think-aloud protocols are well aligned. In the paper-and-pencil measures participants read an average of 3.1% of words incorrectly in connected text, and an average of 1.7% of words incorrectly in the think-aloud text. There were, however, different instructions for the two tasks. The directions for the fluency measure asked participants to “read the passage as accurately as you can at your normal reading speed,” whereas the directions for the think-aloud instructed them to “read the passage as if you were learning the material for a class.” These instructions could have led participants to put special effort

into being more accurate or to focus on their speed in the fluency measure, whereas the instructions for the think-aloud did not focus their efforts on either accuracy or speed, but to read “as if you were learning the material for a class.” However, note that participants were slightly *more* accurate in the think-aloud text, where they were not asked specifically to be accurate.

Results of the Spearman rank correlation analyses provide correlational evidence to corroborate the correlations that underlie the path model: participants who had a high proportion of miscues tended to have a high proportion of misunderstanding vocabulary, tended to less often activate accurate background knowledge at high proportions, and tended to have low free recall scores.

In addition, there was some evidence in the think-aloud protocols to support two relationships not in the model: participants who had more errors in word reading also had difficulty accurately summarizing and drawing inferences, and more often expressed a lack of understanding of the passage. For example Nina read “Meanwhile, Parliament’s effort to get revenue went against the colonists’ belief that they could only be taxed by their own leg-islators. A little confusing.”

Together, these analyses of the product measures, analyses of the process measure, and analyses that coordinate the product and process measure provide converging evidence that, in this non-clinical population, basic decoding is not currently a large source of comprehension problems for high school students, as hypothesized by, e.g., Greene (1998) and Moats (2000). This conclusion is in line with those of Leach et al. (2003) with 4th and 5th grade students, and Buly and Valencia (2002) with students who had failed the Washington State exam for promotion to 5th grade. The results do suggest that fluency plays an important role in 9th grade students’ comprehension, and

that it may contribute to comprehension difficulties for those students who struggle to understand what they read, consistent with Artelt et al.'s (2001) findings in the PISA reading comprehension exam with high school students.

Strategies. Consistent with prior research on high school students (e.g., Reynolds et al., 1990), strategies had an effect on comprehension, but it was a small total effect. Unlike a large body of previous research which did not consider indirect effects, here strategies had a non-significant direct effect on comprehension, but a small but significant indirect effect via inference (e.g., Mathewson, 1989).

Paper-and-pencil strategy items that were difficult for participants included predicting what might follow a passage about a doctor's patients paying for services (40% correct), summarizing a paragraph about the Stamp Act (44% correct), and choosing a sentence that could be deleted from a literary passage (46% correct).

In the think-alouds, participants both accurately and inaccurately enacted a wide range of strategies, especially *summarizing* what was read, *re-reading* text, verbalizing a *feeling of understanding* or *judgment of learning*, *self-questioning*, *taking notes*, or verbalizing that they were *not thinking*. Across all participants, accurate strategies were 22.2% of verbalizations and inaccurate strategies were 11.6%. An example of an accurate SUM+ was "So Britain needed a lot of soldiers to protect all the colonists from the Indians." As an example of an inaccurate *summary*, one participant read, "The delegates saw the need for the colonies to put aside rivalries over land claims and trade in order to meet the common threat." then stated "Uh . . . they found that there might have been an issue so they tried to meet in the middle somewhere." This inaccurate use of strategies was consistent with previous think-aloud research that has coded for accurate and inaccurate use of strategies (e.g., Butler & Winne, 1995; McNamara, 2003).

Results of the Spearman rank correlation analyses provide correlational evidence to corroborate the correlations that underlie the path model: participants who enacted a high proportion of accurate use of strategies tended to have higher free recall scores, but they also verbalized a low proportion of accurate inferences. Participants who enacted a high proportion of inaccurate strategies less often made accurate inferences. There are thus contradictory findings for the relationship between strategies and inference in the think-aloud data.

The think-alouds suggested that participants who frequently used inaccurate strategies (predominantly inaccurate summaries) had a poor understanding of what they read. Like Williams' (1993) learning disabled students, irrelevant or inaccurate prior knowledge often intruded into their summaries, as when Bill state that colonists "felt like they were taxed for no reason," reflecting a present-day attitude towards taxes that was not present in the text.

Together, these analyses of the product measures, analyses of the process measure, and analyses that coordinate the product and process measure provide converging evidence that strategies make an important indirect contribution to 9th grade students' comprehension, and may contribute to comprehension difficulties for those students who struggle to understand what they read.

Comprehension. The large and significant correlation between Gates-MacGinitie comprehension scores and verbal recall scores ($r [44] = .40, p < .05$) suggests that the paper-and-pencil and process measures tap the same underlying abilities. All four significant direct paths from components to comprehension (the direct path from strategies was non-significant) were also present in the think-alouds. Recall, also, that a

number of the inference codes (e.g., LINK) involve activation of background knowledge, but are not included in the background knowledge codes.

In summary, then, all five predictor variables made significant contributions, directly and/or indirectly to reading comprehension, consistent with the large body of published research that was used to build the model. There is converging evidence for the importance of all of the components from product and process measures, as well as from the Spearman rank correlation analyses. These results paint a rich portrait of reading comprehension in 9th grade as a complex process in which many components interact to yield comprehension. In contrast to single-variable studies, these results suggest that all of the components are important and that none of them can be neglected.

Differences Between High- and Low-comprehending Readers

Students with below-median comprehension scores performed significantly lower on every component of comprehension, both those parts of the measures related to the Gates-MacGinitie content and those parts related to the Revolutionary War content used in the think-alouds. Effect sizes were all large, ranging from Cohen's $d = .63$ for word reading errors on the fluency passage to $d = 2.09$ for Gates-MacGinitie vocabulary questions. The pattern of results was identical whether the sample was split at the median or at the 50th percentile (where the mean percentile for the low-comprehending group was 28th percentile and the mean for the high-comprehending group was 76th percentile). Overall, participants had very flat profiles on the component measures; those who had comparatively low scores on one measure tended to have similarly low scores on all of the measures. This finding is consistent with a number of multivariate models of reading comprehension (see, e.g., Carr & Levy, 1990), but stands in contrast to studies of clinical

populations, where uneven profiles have been found (see, e.g., Guthrie, Goldberg, & Finucci, 1972; Snow & Strucker, 2000; but see Sabatini, 2002).

Vocabulary. For vocabulary related to the Gates-MacGinitie, the mean scores for the high-comprehending group were 48% higher than for low-comprehending group. For vocabulary related to the think-aloud text, that difference was 50%. That the largest effect size across all paper-and-pencil measures was for vocabulary is consistent with the path model, supporting the recommendation that vocabulary instruction may benefit these high school students who struggle with comprehension.

In the think-alouds, high comprehenders verbalized a higher proportion of understandings of vocabulary than did low comprehenders (a mean of 5.0% of verbalizations vs. 3.2%); whereas low comprehenders verbalized a higher proportion of misunderstandings of vocabulary than did high comprehenders (3.1% vs. 1.5%). This pattern was consistent with that in the think-alouds from the preliminary study ($n = 14$; Cromley & Azevedo, 2004b).

Background knowledge. For background knowledge related to the Gates-MacGinitie, the mean scores for the high-comprehending group were 47% higher than for low-comprehending group. For background knowledge related to the think-aloud text, that difference was 73% (Cohen's $d = 1.66$ and 1.53 , respectively).

In the think-alouds, high comprehenders verbalized a higher proportion of accurate activation of prior knowledge than did low comprehenders (a mean of 2.5% of verbalizations vs. 1.7%); however, low comprehenders also verbalized a lower proportion of activation of inaccurate or irrelevant prior knowledge than did high comprehenders (.6% vs. .9%). Recall also that activation of background knowledge was also required in order to make certain inferences (i.e., evaluation, knowledge elaboration, and links); high

comprehenders verbalized these types of accurate inferences more than twice as often as did low comprehenders (23.6% vs. 11.4%). This pattern was consistent with that in the think-alouds from the preliminary study ($n = 14$; Cromley & Azevedo, 2004b).

Inference. For inferences related to the Gates-MacGinitie, the mean scores for the high-comprehending group were 74% higher than for low-comprehending group. For inferences related to the think-aloud text, that difference was 108% (Cohen's $d = 1.76$ and 1.71, respectively).

In the think-alouds, high comprehenders verbalized a higher proportion of accurate inferences than did low comprehenders (a mean of 30.8% of verbalizations vs. 16.7%); however, low comprehenders also verbalized fewer inaccurate inferences than did high comprehenders (2.6% vs. 3.7%). This pattern was consistent with that in the think-alouds from the preliminary study ($n = 14$; Cromley & Azevedo, 2004b).

Word reading. For word reading accuracy, the mean score on real word reading for the high-comprehending group was 8% higher than the low-comprehending group (Cohen's $d = 1.22$), and for pseudoword reading it was 17% higher (Cohen's $d = .86$). For word reading fluency, the mean speed in words per minute for the high-comprehending group was 20% faster than the low-comprehending group (Cohen's $d = 1.06$).

In the think-alouds, low comprehenders verbalized a higher proportion of word reading errors, omissions, and self-corrections than did high comprehenders (a mean of 39.3% of verbalizations vs. 23.5%). This pattern was consistent with that in the think-alouds from the preliminary study ($n = 14$; Cromley & Azevedo, 2004b).

Strategies. For strategies related to the Gates-MacGinitie, the mean scores for the high-comprehending group were 94% higher than for low-comprehending group. For

strategies related to the think-aloud text, that difference was 44% (Cohen's $d = 1.76$ and $.88$, respectively).

In the think-alouds, high comprehenders verbalized a higher proportion of accurate uses of strategies than did low comprehenders (a mean of 24.5% of verbalizations vs. 19.1%); whereas low comprehenders verbalized a higher proportion of inaccurate uses of strategies than did high comprehenders (13.3% vs. 7.6%). This pattern was consistent with that in the think-alouds from the preliminary study ($n = 14$; Cromley & Azevedo, 2004b).

In summary, both the paper-and-pencil measures and the think-aloud protocols suggest large differences between low-and high-comprehending students across all of the components of reading comprehension. However, the largest differences in the paper-and-pencil measures were on vocabulary, whereas the largest differences in the think-alouds were on inferences. This may be due to the way each variable is operationalized in the paper-and-pencil measures, the effect of test-taking strategies on performance on the paper-and-pencil measures, the way each variable is coded in the think-aloud coding scheme, what participants are or are not willing to verbalize in the think-alouds, or some combination(s) of these.

Comprehension. For literal questions on the Gates-MacGinitie, the mean scores for the high-comprehending group were 54% higher than for low-comprehending group. For inference questions, that difference was 59% (Cohen's $d = 2.75$ and 2.77 , respectively).

On the think-alouds, summing up across all of the coding categories, 62.8% of high comprehenders' verbalizations were accurate and 37.2% were inaccurate, whereas 40.7% of low comprehenders' verbalizations were accurate and 59.3% were inaccurate.

Proportion of accurate verbalizations had a Spearman rank correlation with comprehension scores of $r_s [44] = .52, p < .05$. Participants who scored high on the Gates-MacGinitie tended to verbalize a higher proportion of accurate codes, whereas those with lower Gates scores tended to verbalize a lower proportion of accurate codes (and therefore a higher proportion of inaccurate codes). This pattern was consistent with that in the think-alouds from the preliminary study ($n = 14$; Cromley & Azevedo, 2004b).

The results are consistent with several recent studies providing evidence that students can fall “off track” at many points in their reading development, not only in word reading (Buly & Valencia, 2002; Leach et al., 2003; Spear-Swerling & Sternberg, 1996). Considered developmentally, students in this sample have mastered the prerequisites of word reading and reading comprehension strategies better than they have inferences, but where they really lag behind is in background knowledge, and especially in academic vocabulary.

Taken together, these results suggest that reading comprehension is a highly complex cognitive task that involves many components which act both separately and together. For example, background knowledge is needed not only for literal comprehension, but is also needed in order to draw inferences (which is at the heart of Kintsch’s [1998] CI model). Variables such as background knowledge and vocabulary that have effects on intermediate variables, make a larger contribution to comprehension because they have both direct and indirect effects. This means that students who have strong skills in these areas get an extra benefit from both direct and indirect contributions to comprehension, and that students with weak skills are at an extra disadvantage. For example, even with good inference skills, a student who lacks background knowledge may not be able to make sense of a passage. Hypothetically, this could make sense of the

contradictory evidence for Matthew effects in vocabulary and comprehension. If the gap in vocabulary knowledge between lower- and higher-comprehending students remains stable over time, the gap in comprehension itself could nonetheless continue to grow.

The results also suggest that high school students who struggle with reading comprehension struggle with all of the components at once. Results of the path analysis suggest that vocabulary and background knowledge, two of the most distal variables, make the largest contribution to comprehension. This should not be interpreted to mean that students do not need inference, strategy, or even word reading help, but that help with these components would be expected to yield smaller rewards, unless instruction in those components also increased vocabulary and background knowledge.

Limitations

This study considered students from only one school, in one grade, and at one time; however, the sample was relatively large and was very diverse. While the target school is a multi-ethnic school in a suburb of a major city, the sample cannot represent all high school students. The measures tapped only school-based reading texts and tasks, but not the whole range of reading activities that adolescents engage in (Ivey, 1999; Moje, 2001). The think-aloud portion of the study used only one text from only one domain (social studies). In addition, the effect of genre or text structure is not considered; students might show different levels of proficiency in different text types.

As Snow (2002) points out, no existing reading comprehension measures are based on theories of reading comprehension. Findings from this study are specific to the measures that were used. While the measures show good reliability with this sample (except for strategy questions related to the Revolutionary War), their reliability was not perfect, and was not the same from measure to measure, which complicates interpretation

of the path loadings. There is substantial evidence for validity of the measures, but evidence from the paper-and-pencil measures and think-alouds was not totally consistent. As with all measures, scores could be confounded with other unmeasured variables such as attention—students in the non-honors classes had to be reminded to focus on the measures, whereas students in the honors classes did not need to be reminded.

While this study considered the relative contribution of a set of five predictor variables, that set of variables was limited by prior research. For example, what role(s) might motivation, attention, working memory, or processing speed play in comprehension for 9th grade students? In addition, the interrelationships posited in the four competing models were based on prior research, but what important interrelationships have not been previously studied (e.g., the effect of vocabulary knowledge on accurate use of strategies)? This study was restricted to confirmatory analyses, in order to avoid the risk of over-fitting the model and capitalizing on chance. Now that the model has been validated, future studies might explore these “missing” paths or drop paths that were non-significant across both the current study and the preliminary study.

While the paper-and-pencil measures show relatively “flat” profiles across the component measures, the think-aloud protocols hint at some possible individual differences among students. Specifically, participants occasionally verbalized errors in, e.g., vocabulary, followed by accurate use of strategies or accurate inferences. For example, Peter showed a misunderstanding of *smuggled*, “one thing that was smuggled was- that was stolen or- was molasses,” but followed this with an accurate summary “and they used that to make rum in New England.” This sequence of verbalizations should not occur, according to the model. It is possible, therefore, that some participants are using

some sort of compensatory strategy or strategies that enable them to comprehend (see, e.g., Walczyk, 2000). These possible individual differences should be explored in future research.

It is possible that paths with strong empirical support (e.g., from strategies to comprehension) are not significant because of the particular sample drawn. Omitting relevant variables is also a threat to interpreting path models—a variable could be identified as making a large contribution, but which is really the product of some other omitted variable (e.g., working memory) that should have been included. Future research should consider larger samples and more indicators for each component so that a structural equation model can be tested. There is also a need for multi-year developmental studies of reading comprehension and its components (NICHD, 2002), rather than the “one shot” picture collected here.

Covariance structure modeling, of which path analysis is a subtype, can never “prove” a model—the best-fitting model is only one of many possible models that could explain the same data (the same variance/covariance matrix). However, this design has the advantage that multiple paths in theoretically-driven models can be tested simultaneously, rather than conducting a separate experiment for each direct effect.

Implications

These findings have implications for theories of reading comprehension, for future reading comprehension research, and for teachers and administrators who are responsible for educating high school students.

Theoretical contributions. The results of this research add to our understanding of the roles of different components in reading comprehension. This study both validates and refines the Inferential Mediation model, a new model of reading comprehension

(Cromley & Azevedo, 2004b). The path analysis has shown the direct and indirect contribution of each component to reading comprehension in 9th grade. Rather than the largely indirect paths suggested by other theories, the IM model suggests that direct effects are most important. This theoretical model also has practical implications, e.g., for explaining why certain intervention programs produce the results that they do, and for choosing interventions.

For future research.

Measures. This study has refined two measures—the background knowledge and inference measures—from the preliminary study (Cromley & Azevedo, 2004), and added a new strategy use measure that is both reliable and valid with this 9th grade sample. These measures may be useful for other research, especially for studies using the Gates-MacGinitie Level 7/9, Form S comprehension subtest (MacGinitie et al., 2001).

Methodological contributions. This study combines multi-component product (test scores) and process (think-aloud) data to investigate reading comprehension. Existing multi-method studies in reading suffer from methodological problems such as calculating Pearson correlations between scale data and frequency data. The study makes two methodological improvements to coordinating product (scale data) and process data (frequency data). First, using Spearman rank correlations between the relevant components and variables from the think-aloud data avoids violating assumptions of the Pearson correlation, and yielded significant results as in Chung and colleagues' (2002) study of a hypermedia assessment system. Second, the co-occurrence of pairs of variables in the protocols was investigated when significant direct and indirect paths were found in the path analysis, adapting a method suggested by Winne and colleagues (Winne et al.,

2002). This methodology allows for converging data about comprehension from two different sources.

Practical contributions. The results of the path analysis can point in some specific direction(s) for designing future interventions that have the most potential to improve young adolescents' reading. First, the results can point to the most influential predictor variables (combining both indirect and direct effects). Second, the results can point to the need to address more than one variable at a time (e.g., vocabulary and background knowledge). The finding across both the preliminary study and the dissertation that vocabulary made the largest total contribution to comprehension provides a strong direction for teachers and administrators who wonder where to begin to address high school students' reading comprehension problems.

The large contribution to comprehension by vocabulary suggests that vocabulary instruction might be the most beneficial approach for 9th grade students who struggle with comprehension (assuming that such individuals show vocabulary difficulties and do not have excessively slow or inaccurate word reading). It is unclear from this study, however, whether it is the meanings of root words that is the largest obstacle for 9th grade students, whether it is assembling prefixes or suffixes and root words (e.g., the “-en/-an” suffix as in “Serbia” and “Serbian” or “Germany” and “German”), or some combination of difficulties with root words, prefixes, and suffixes.

APPENDIX A

*Comprehension Strategy Instruction Studies Showing Effects on Reading Comprehension
With Middle- and High-School Students, Listed in Order by Strategy*

Appendix A

Middle School and High School Comprehension Strategy Instruction Studies Showing Effects on Reading Comprehension, Listed in

Order by Strategy

Title	Participants	Strategy	Domain	Posttest	Results
Alvermann, D. (1982). Restructuring text facilitates written recall of main ideas. <i>Journal of Reading</i> , 81(6), 754-758.	30 10 th grade 5 th -7 th stanines, Stanford	Graphic organizer (researcher-developed)	Biology (whales)	Written recall, top-level structure	Strategy group significantly higher on both amount recalled and top-level structure
Alvermann, D. E. (1981). The compensatory effect of graphic organizers on descriptive text. <i>Journal of Educational Research</i> , 75 (1), 44-48.	114 10 th grade, Regents & non-Regents	Graphic organizer	Biology (loss of body water)	Written recall	GO signif. better for both hi & lo comprehenders, but only for description text
Berkowitz, S. J. (1986). Effects of instruction in text organization on sixth-grade students' memory for expository reading. <i>Reading Research Quarterly</i> , 21(2), 161-178.	99 6 th grade low and high (Gates-MacGinitie)	Graphic organizer (Active vs. Passive) vs. C1 (question answering) C2 (reread)	Social studies (textbook)	Written free recall Short answer Immediate, 2-week delay & transfer	GOA = GOP > C1 >C2 Short answer main ideas and total recall
Alvermann, D. E., & Hynd, C. E. (1995). Effects of interactive discussion and text type on learning counterintuitive science concepts. <i>Journal of Educational Research</i> , 88 (3), 146-154.	86 9 th grade general science, 60% African-American	Graphic organizer (discussion web) vs. worksheet	Physics (projectile motion)	Pre- & Post: 10-item 3-choice knowledge, 10-item vocabulary, 10-item misconceptions, & application (diagram). Post only: cued recall	Miscon GO>W=C Applic (GO>W)=C

Title	Participants	Strategy	Domain	Posttest	Results
Alvermann, D. (1988). Effects of spontaneous and induced lookbacks on self-perceived high- and low-ability comprehenders. <i>Journal of Educational Research</i> , 81(6), 325-331.	64 10 th high: 99 th -83 rd %ile low: 42 nd – 18 th %ile on TAP	Graphic organizer (researcher-developed)	Social studies (Louis XIV, diffusion of Western culture)	20-item free-response x 2 passages: familiar and unfamiliar	Low-verbal students in strategy group significantly higher.
Stensvold, M. S., & Wilson, J. T. (1990). The interaction of verbal ability with concept mapping in learning from a chemistry laboratory activity. <i>Science Education</i> , 74(4), 473-480.	104 9 th grade science students (higher than average on Iowa)	Graphic organizer (student developed)	Chemistry (lab—chemical reactions)	10 fact + 15 concept + 8 application/ analysis questions, concept map words, links [Tment & control equal on Iowa]	Concept map-ing benefited low- vocabulary students, only on comprehen- sion test
Bean, T. W., Singer, H., Sorter, J., & Frazee, C. (1986). The effect of metacognitive instruction in outlining and graphic organizer construction on students' comprehension in a tenth-grade world history class. <i>Journal of Reading Behavior</i> , 18(2), 153-169.	72 10 th grade honors World History (Tment 1 participated in Bean et al., 1983)	Graphic organizer (student-developed)	World History	15 item x 6 teacher- devised multiple-choice quizzes, delayed written recall	Strategy group signif. better on 6 th of 6 quizzes + quality of written recall
Bean, T. W., Singer, H., & Sorter, J. (1987). Acquisition of hierarchically organized knowledge and prediction of events in world history. <i>Reading Research and Instruction</i> , 26(2), 99-114.	47 10 th grade honors World History (average and above on STEP)	Graphic organizer (student-developed) + prediction vs. Outline + recitation	World History (American & French revolutions)	5 textbook multiple- choice quizzes, transfer tests: essay and 10 multiple choice prediction questions	Strategy group signif. better on essay transfer test only
Gordon, C. J., & Rennie, B. J. (1987). Restructuring content schemata: An intervention study. <i>Reading Research and Instruction</i> , 26(3), 162-188.	23 5 th grade	Graphic organizer vs. C1 (no-instruction summary) vs. C2	Biology (altruism in animals) and related narrative	Short answer and Yes/No questions	GO > C1 = C2

Title	Participants	Strategy	Domain	Posttest	Results
Gallini, J. K., Spires, H. A., Terry, S., & Gleaton, J. (1993). The influence of macro and micro-level cognitive strategies training on text learning. <i>Journal of Research and Development in Education</i> , 26(3), 164-178.	66 High school remedial (no grade spec.)	Graphic organizer vs. anaphora	Biology	10 multiple choice questions x 6 assmts + summary; 3 wks delayed	Graphic organizer group signif. better overall, both tests.
Armbruster, B. B., Anderson, T. H., & Meyer, J. L. (1991). Improving content area reading using instructional graphics. <i>Reading Research Quarterly</i> , 26(4), 393-416.	365 164 4 th grade 201 5 th grade	Graphic organizer vs. Control	Social studies (textbook)	Short answer	GO > C for 5 th grade only
Darch, C. B., Carnine, D. W., & Kameenui, E. J. (1986). The role of graphic organizers and social structure in content area instruction. <i>Journal of Reading Behavior</i> , 18(4), 275-295.	84 6 th grade	Graphic organizer vs. Directed reading vs. SQ3R	Social studies	Short answer and fill-in-the blank Immediate and 1-day delayed	GO = SQ3R > DR immediate SQ3R > GO > DR delayed
Hafner, L. E., & Palmer, B. C. (1980). The differential effects of three methods of teaching on the reading comprehension and vocabulary of ninth grade students. <i>Journal of Educational Research</i> , 74(1), 34-37.	80 9 th heterogeneous group	Logical relations (MRM, LOGNAR, vocab, & control)	Varied reading passages?	Nelson-Denny vocabulary & Davis comprehension	MRM = LOG > V = C

Title	Participants	Strategy	Domain	Posttest	Results
Sjostrom, C. L., & Hare, V. C. (1984). Teaching high school students to identify main ideas in expository text. <i>Journal of Educational Research</i> , 78(2), 114-118.	19 9 th & 10 th Voluntary academic enrichment program Hispanic & African-American Upper 25% on Davis Reading Test	Main idea	Expository text	Experimenter-developed science summarizing test & Davis Reading Test	Strategy group signif. better on experimenter-developed test only
Alfassi, M. (2004). Reading to learn: Effects of combined strategy instruction on high school students. <i>Journal of Educational Research</i> , 97(4), 171-184.	Study 1: 49 9 th grade students from 2 mixed-ability language arts classes	Study 1: Multiple (reciprocal teaching) vs. Regular class	Various expository	Experimenter-developed test & Gates-MacGinitie	Strategy group signif. better on both
Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. <i>Cognition and Instruction</i> , 2, 117-175.	37 7 th grade 24 struggling readers (good decoding, poor comprehension) 13 non-struggling	Multiple (reciprocal teaching) vs. Locating information vs. Daily test only vs. Pretest-posttest control	Various expository	Researcher-developed daily tests, generalization probes (quizzes), 4 transfer tests (summarizing, predicting questions, detecting incongruities, rating importance) and Gates-MacGinitie vocabulary and comprehension	RT group significantly better than control on all researcher-developed measures except predicting questions and rating importance.
Taylor, B. T., & Frye, B. J. (1992). Comprehension strategy instruction in the middle grades. <i>Reading Research and Instruction</i> , 32(1), 39-48.	150 5 th & 6 th grades average and above (teacher rating and test scores)	Multiple (reciprocal teaching) vs. Control (basal)	Social studies	Summarize Generate questions Short answer	S > C for summaries only

Title	Participants	Strategy	Domain	Posttest	Results
Alfassi, M. (1998). Reading for meaning: The efficacy of reciprocal teaching in fostering reading comprehension in high school students in remedial classes. <i>American Educational Research Journal</i> , 35(2), 309-332.	75 9 th remedial	Multiple—reciprocal teaching	Varied (remedial reading)	Daily: Reading w/10 free-response comprehension questions. Posttest: GatesMacGinitie vocabulary and comprehension.	Strategy group significantly higher on researcher-designed measures, not on standardized
Guthrie, J. T., Van Meter, P., Hancock, G. R., Alao, S., Anderson, E., & McCann, A. (1998). Does concept-oriented reading instruction increase strategy use and conceptual learning from text? <i>Journal of Educational Psychology</i> , 90(2), 261-278.	172 students 23 rd -42 nd %ile 90 3 rd grade & 82 5 th grade	Multiple— Concept-Oriented Reading Instruction (CORI)	Science	Strategy use, relevance, reasons, notes, new conceptual knowledge, draw, write, text comprehension, story comprehension, informational text comprehension	CORI > Control on strategy use, new conceptual knowledge, conceptual transfer (Gr. 5), story comprehension
Guthrie, J.T., Anderson, E., Alao, S., & Rinehart, J. (1999). Influences of concept-oriented reading instruction on strategy use and conceptual learning from text. <i>Elementary School Journal</i> , 99, 343-366.	229 students 30 th %ile 123 3 rd grade & 106 5 th grade	Multiple— Concept-Oriented Reading Instruction (CORI)	Science	Strategy use, conceptual knowledge (draw and write), transfer, story comprehension, informational text comprehension	CORI > Control on most texts, most grades for all measures
Guthrie, J.T., Wigfield, A., & VonSecker, C. (2000). Effects of integrated strategy instruction on motivation and strategy use in reading. <i>Journal of Educational Psychology</i> , 92(2), 331-341.	162 students (2 Chapter I schools) 74 3 rd grade & 88 5 th grade	Multiple— Concept-Oriented Reading Instruction (CORI)	Science	Curiosity, involvement, strategy use, recognition, competition	CORI > Control on curiosity and strategy use

Title	Participants	Strategy	Domain	Posttest	Results
Leon, J. A., & Carretero, M. (1995). Intervention in comprehension and memory strategies: Knowledge and use of text structure. <i>Learning and Instruction</i> , 5(3), 203-220.	72 9 th and 10 th grade 36 low 36 high Composite of standardized test, teacher and summary writing	Multiple (text structure, summarizing, graphic organizer) vs. C1 (same materials, encouraged to use known strategies) vs. C2	Social studies (textbook)	Free recall immediate delayed main ideas details structures organization of recall	M > C1 = C2 on immediate structures
Groller, K. L., Kender, J. P., & Honeyman, D. S. (1991). Does instruction on metacognitive strategies help high school students use advance organizers? <i>Journal of Reading</i> , 34, 470-475.	45 11 th grade, high reading (>80 th %ile on MAT)	Multiple—Graphic organizer + Metacognitive	Philosophy (Descartes)	Short answer (NO PRETEST)	AA + M > AA = control
Mason, L. (2004). Explicit self-regulated strategy development versus reciprocal questioning: Effects on expository reading comprehension among struggling readers. <i>Journal of Educational Psychology</i> , 96(2), 283-296.	32 5 th grade struggling comprehenders	Multiple—adapted KWL (called TWA) vs. Re-Quest	Science and social studies	Oral and written comprehension	TWA > RQ for oral but not written comprehension
Prince, A. T., & Mancus, D. S. (1987). Enriching comprehension; A schema altered basal lesson. <i>Reading Research and Instruction</i> , 27(1), 45-54.	45 1 st -5 th grades, 5 th grade reading 6 th -8 th grade level material	Post-reading discussion vs. Control	Basal narrative and expository	Short answer	P > C (did not separate results by grade)

Title	Participants	Strategy	Domain	Posttest	Results
Mathewson, G. C. (1989). Effects of if-then usage upon urban students' inference generation during American history reading. In S. McCormick & J. Zutell (Eds.), <i>Cognitive and social perspectives for literacy research and instructions, 38th Yearbook of the National Reading Conference</i> (pp. 331-338). Chicago, IL: National Reading Conference.	24 11 th grade Stanford %iles 67 72 (AP) 52, 48, 41	Prediction vs. Note-taking vs. Summarizing	American history	Predictions from paragraph; repeated 1 wk later	Prediction = Note-taking Sig > Summarizing for amount written; Prediction Sig > Note-taking = Summarizing for causal links & relevance
Andre, M. E. D. A., & Anderson, T. H. (1978-79). The development and evaluation of a self-questioning study technique. <i>Reading Research Quarterly, 14</i> (4), 605-623.	St 1: 29 St 2: 81 Study 1: 12 th Study 2: 11 th -12 th	Question generation	Psychology	Study 1: 20-item free-response x 2 passages: main idea & detail questions Study 2: 8-item free-response x 2 passages: 4 main idea & 4 detail questions. Students also write own questions	Study 1: Low-verbal students in strategy group significantly higher. Study 2: Strategy group significantly higher, write better questions
Frase, L. T., & Schwartz, B. J. (1975). Effect of question production and answering on prose recall. <i>Journal of Educational Psychology, 67</i> (5), 628-635.	48 11 th -12 th grade	Question generation	Social studies (biography)	90 tape-recorded short-answer questions	Questioning = answering sig. > solo phase; if student made up a question seen later on the test, got more correct; more time spent studying in questioning/ answering phases

Title	Participants	Strategy	Domain	Posttest	Results
Singer, H., & Donlan, D. (1982). Active comprehension: Problem-solving schema with question generation for comprehension of complex short stories. <i>Reading Research Quarterly</i> , 17(2), 166-186.	27 11 th grade	Question generation	Literature (difficult short stories)	10 item multiple choice x 6 quizzes	Strategy group signif. better
Wong, B. Y. L., & Jones, W. (1982). Increasing metacomprehension in learning disabled and normally achieving students through self-questioning training. <i>Learning Disability Quarterly</i> , 5, 228-240.	120 8 th -9 th grade LD (n = 60, 5.6 GE on Nelson) 6 th grade non-LD (n = 60, 6.7 GE)	Question generation	Passages from standard test lessons	14 paragraphs, underline main idea, formulate question; 5 x 5-paragraph passages, formulate questions, & comprehension questions	Only LD group significantly better on all measures
Dreher, M. J., & Gambrell, L. B. (1985). Teaching children to use a self-questioning strategy for studying expository prose. <i>Reading Improvement</i> , 22, 2-7.	60 6 th grade boys	Question generation (Direct vs. Implicit) vs. Control (reread)	Not specified	Summary writing (main ideas and supporting details), immediate and delayed	QD > QI for main ideas only on delayed test
Nolte, R. Y. & Singer, H. (1985). Active comprehension: Teaching a process of reading comprehension and its effects on reading achievement. <i>The Reading Teacher</i> , 39, 24-31.	40 5 th grade	Question generation vs. Control (pre-teach vocabulary)	Narrative	Comprehension questions	QG > C
Helfeldt, J. P., & Lalik, R. (1976). Reciprocal student-teacher questioning. <i>The Reading Teacher</i> , 30, 283-287.	22 5 th grade	Question generation vs. Teacher questions	Not specified	Van Wageningen Analytical Reading Scale	QG > TQ

Title	Participants	Strategy	Domain	Posttest	Results
Davey, B., & McBride, S. (1986). Effects of question-generation training on reading comprehension. <i>Journal of Educational Psychology</i> , 78(4), 256-262.	120 6 th grade Native English speakers	Question training (QT) vs. No-Question Control (NQC) vs. Question Generation (GP) vs. Inference Practice (IP), Literal question practice (LP)	Expository (profiles of people in challenging jobs)	Short answer literal inferential Question generation	GT = GP > NQC = IP for literal QT > all for inferential and question generation
Lysynchuk, L. M., Pressley, M., & Vye, N. J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. <i>The Elementary School Journal</i> , 90(5), 469-484.	72 36 4 th <50%ile MAT 36 7 th <50%ile Gates-MacGinitie	Multiple—Reciprocal teaching vs. Control (small group)	Expository	Oral free recall Questions explicit implicit Standardized comprehension Standardized vocabulary	RT > C for all except Standardized vocabulary
Symons, S., McLatchy-Gaudet, H., Stone, T. D., & Reynolds, P. L. (2001). Strategy instruction for elementary students searching informational text. <i>Scientific Studies of Reading</i> , 5(1), 1-33.	Experiment 1: 180 60 3 rd grade 60 4 th grade 60 5 th grade	Search (SLI vs. SLI + monitoring) vs. Control	Social studies (history—Wright Brothers)	Time to answer 3 questions by searching	SLI + M > SLI > C
Carnine, D. W., Kameenui, E. J., & Woolfson, N. (1982). Training of textual dimensions related to text-based inferences. <i>Journal of Reading Behavior</i> , 14(3), 335-340.	36 5 th grade	Summarize vs. C1 (feedback) vs. C2	3 stories	Short answer	S > C1 = C2
Rinehart, S. D., Stahl, S. A., & Erickson, L. G. (1986). Some effects of summarization training on reading and studying. <i>Reading Research Quarterly</i> , 21(4), 422-438.	70 6 th grade above average on Gates-MacGinitie	Summarize vs. Control	Social studies (textbook)	Summary Recall major ideas minor	S > C summary and major ideas

Title	Participants	Strategy	Domain	Posttest	Results
Bean, T. W., Singer, H., Sorter, J., & Frazee, C. (1983). Acquisition of summarization rules as a basis for question generation in learning from expository text at the high school level. In J. A. Niles (Ed.), <i>Searches for meaning in reading/language processing and instruction</i> (32 nd Yearbook of the National Reading Conference (pp. 43-49). Rochester, NY: National Reading Conference.	58 10 th grade honors World History avg. & above on STEP	Summarizing	Greece and Rome	33 item x 8 teacher-devised multiple-choice quizzes, summary of 300-word text	Strategy group signif. shorter summary, but same quality, same on quizzes
Bean, T. W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. <i>Journal of Reading Behavior</i> , 16(4), 297-306.	60 6 th grade	Summarizing (DI vs. Implicit) v. Control (advice to summarize)	Basal	Summary Multiple-choice comprehension	Sum DI = Sum I > C
Armbruster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? <i>Reading Research Quarterly</i> , 22(3), 331-346.	82 5 th grade non-remedial	Summarizing v. Control (text and questions)	Social studies (textbook)	Summary (essay) Short answer	S > C on summary only
Baumann, J. F. (1984). The effectiveness of a direct instruction paradigm for teaching main idea comprehension. <i>Reading Research Quarterly</i> , 20(1), 93-115.	61 6 th grade low, middle, and high on Metropolitan Achievement Test (MAT)	Summarizing vs. Basal treatment v. Control (vocabulary)	Social studies	Multiple-choice comprehension (main idea and details) Summarizing Written free recall	T > B > C on MC and Sum Better readers benefited more

Title	Participants	Strategy	Domain	Posttest	Results
Taylor, B. M., & Beach, R. W. (1984). The effects of text structure instruction on middle-grade students' comprehension and production of expository text. <i>Reading Research Quarterly, 19</i> (2), 134-146.	114 7 th grade (8.5 on California Achievement Test)	Summarizing vs. Control (text and questions v. no treatment)	Social studies (textbook)	Written free recall Short answer	S > C1 = C2 for recall S = C1 > C2 for short ans.
Short, J., & Ryan, E. B. (1984). Metacognitive differences between skilled and less skilled readers: Remediating deficits through story grammar and attribution training. <i>Journal of Educational Psychology, 76</i> (2), 225-235.	42 4 th grade	Story structure with attribution vs. Story structure only vs. Attribution only	Short stories	Free recall, short-answer questions, delayed error detection, Reading Concept Inventory	All measures: SA = S > A
Fitzgerald, J., & Spiegel, D. L. (1983). Enhancing children's reading comprehension through instruction in narrative structure. <i>Journal of Reading Behavior, 15</i> (2), 1-17.	19 4 th grade below average on both CAT and story grammar tasks	Story structure vs. dictionary instruction	Short stories	Story production, scrambled story, comprehension questions (no lookback)	SS > DI on all measures
Greenewald, M. J., & Rossing, R. L. (1986). Short-term and long-term effects of story grammar and self-monitoring training on children's story comprehension. In J. A. Niles & R. V. Lalik (Eds.), <i>Solving problems in literacy: Learners, teachers, and researchers. 35th yearbook of the National Reading Conference</i> (pp.210-213). Chicago: National Reading Conference.	22 4 th grade	Story structure vs. Control (basal)	African folktales	Recall questions, written free recall (used on different stories); immediate and 4-week delayed	Immediate: SS > C on recall questions Delayed: SS > C on both

Title	Participants	Strategy	Domain	Posttest	Results
Dimino, J., Gersten, R., Carnine, D., & Blake, G. (1990). Story grammar: An approach for promoting at-risk secondary students' comprehension of literature. <i>Elementary School Journal</i> , 91(1), 19-32.	32 9 th grade	Story structure	Literature	5-7 story grammar questions + 6 factual Qs + summary writing x 2 stories 2 days after; 2 weeks after	Strategy group signif. better on all measures, all times. Students w/low prior story grammar increased sig., those w/hi prior, no increase

APPENDIX B

*Strategies Enacted by Middle and High School Readers in Think-Aloud Studies, Listed
Alphabetically by Author*

Appendix B

Strategies Enacted by Middle and High School Readers in Think-Aloud Studies, Listed Alphabetically by Author

Citation	Type of text, topic	Type of reader(s)									
			Evaluate	Hypothesize/ Predict	Monitor	Reread	Relate to prior knowledge	Summarize/ paraphrase	Text structure	Visualize	
Afflerbach, P. (1990). The influence of prior knowledge and text genre on readers' prediction strategies. <i>Journal of Reading Behavior</i> , 22(2), 131-148.	3 essays and 2 short stories	10 graduate students and 5 high school students from a gifted and talented program		X	X			X			
Christopherson, S. L., Schulz, C. B., & Waern, Y. (1981). The effect of two contextual conditions on recall of a reading passage and on thought processes in reading. <i>Journal of Reading</i> , 23, 573-578.	Bransford & Johnson's (1972) laundry text— with title Without title	High school, grade(s) not specified						X	X		
Fehrenbach, C. (1991). Gifted/average readers: Do they use the same reading strategies? <i>Gifted Child Quarterly</i> , 35(3), 125-127.	Narrative	8 th , 10 th & 12 th grades, 30 gifted (95 th percentile on ITBS comprehension) and 30 average (40 th -60 th percentile on ITBS)	G > P				X	G > P		G > P	G > P

Citation	Type of text, topic	Type of reader(s)								
			Evaluate	Hypothesize/ Predict	Monitor	Reread	Relate to prior knowledge	Summarize/ paraphrase	Text structure	Visualize
Harker, W. J. (1994). "Plain sense" and "poetic significance": Tenth-grade readers reading two poems. <i>Poetics</i> , 22, 199-218.	Two poems, one more abstract, one more literal	15 10 th grade students identified by their teachers as capable in literary interpretation	X		X		X			
Kletzien, S. B. (1991). Strategy use by good and poor comprehenders reading expository text of differing levels. <i>Reading Research Quarterly</i> , 26(1), 67-86.	Expository, frustration level only	24 10 th & 11 th grades Good (>70 th %ile on California Test of Basic Skills) & poor (<50 th %ile)					G > P	G > P	G > P	
Kletzien, S. B. (1992). Proficient and less-proficient comprehenders' strategy use for different top-level structures. <i>Journal of Reading Behavior</i> , 24(2), 191-215.	Collection text Causation text	10 th & 11 th grades Good (>75 th %ile on California Test of Basic Skills) & poor (<50 th %ile)					X	X		X
Meyers, J., Lytle, S., Palladino, D., Devenpeck, G., & Green, M. (1990). Think-aloud protocol analysis: An investigation of reading comprehension strategies in fourth- and fifth-grade students. <i>Journal of Psychoeducational Assessment</i> , 8, 112-127.	Narrative	4 th and 5 th grade average readers		X	X	X	X	X		X

Citation	Type of text, topic	Type of reader(s)									
			Evaluate	Hypothesize/ Predict	Monitor	Reread	Relate to prior knowledge	Summarize/ paraphrase	Text structure	Visualize	
Neuman, S. B. (1990). Assessing children's inferencing strategies. In J. Zutell & J. McCormick (Eds.), <i>Literacy theory and research: Analyses from multiple paradigms</i> (pp. 267-274). Thirty-ninth yearbook of the National Reading Conference. Chicago, IL: National Reading Conference.	Two stories from a children's mystery series.	42 21 low-achieving and 21 high-achieving 5 th grade students (measured with the Metropolitan Achievement Test)					P > G				
Olshavsky, J. E. (1976-77). Reading as problem solving: An investigation of strategies. <i>Reading Research Quarterly</i> , 12(4), 654-764.	Short story	24 10 th grade good & poor		G > P			G > P	G > P			
Phillips, L. M. (1988). Young readers' inference strategies in reading comprehension. <i>Cognition and Instruction</i> , 5(3), 193-222.	3 passages about either a high-familiarity or a low-familiarity topic	40 low-proficiency and 40 high-proficiency 6 th grade students						X			
Pritchard, R. (1990). The effects of cultural schemata on reading processing strategies. <i>Reading Research Quarterly</i> , 25(4), 273-295.	Letter about a funeral—same for culturally familiar and unfamiliar	9 th grade Average (teacher nomination)					X	X	X		X

Citation	Type of text, topic	Type of reader(s)								
			Evaluate	Hypothesize/ Predict	Monitor	Reread	Relate to prior knowledge	Summarize/ paraphrase	Text structure	Visualize
Smith, M. W. (1991). Constructing meaning from text: An analysis of ninth-grade reader responses. <i>Journal of Educational Research</i> , 84(5), 263-271.	Short stories	9 th grade good & poor	G > P		G > P			G > P	G > P	

Note: X indicates the strategy was used. G > P means good readers used this strategy more than did poor readers. Note that because of previously-discussed data analysis problems with the studies, no statistical significance is reported.

APPENDIX C

*Preliminary Study Parent Cover Letter, Parental Consent Form, Student Assent Form,
and Student Background Information Sheet*



Attachment A—Parental permission form

LETTER TO PARENTS

September 16, 2002

Dear Parent/Guardian:

With your permission, we would like to include your child in a study of reading processes used by high school students. The study, which will be conducted at [name of high school], has been approved by University of Maryland's Institutional Review Board (IRB) and your child's teacher, [name of teacher]. We expect the study to be completed by the end of June 2003.

The purpose of this study is to examine high school students' reading processes using standardized measures and by asking some children to "think aloud" while reading. Very little is known about reading processes in high school students, and we hope to learn more from students at various levels of reading proficiency. Your child's teacher may approve this activity for community service credit.

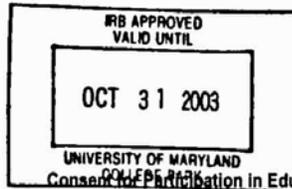
The main method we will use in the study is to complete brief word reading, reading comprehension, reading strategy, background knowledge, and vocabulary measures. In addition, one-half of the students will be selected to "think aloud" while reading a short passage from a high school social studies textbook and recall the passage, all of which we will audiotape. There will be one group session of approximately 90 minutes for all students and a second individual session of 45 minutes for the one-half of the students who participate in the "think-aloud" portion of the study. The study will be conducted at the school, either during class time or after school. We will also collect any 8th or 9th grade reading test scores on file for your child. We will then conduct a detailed analysis of the measures and audiotapes.

We will take a number of steps to ensure your child's privacy and anonymity. For example, no individuals' names will be used in any written or oral presentations resulting from the study and only a limited number of professional researchers will have access to the original tapes and in-class tests. Each student's participation will be fully voluntary. We will describe the general purpose of the study to each student so that each student can decide whether to participate and s/he may withdraw from the study at any time.

We hope that you will consent to your child's participation in this study that may one day contribute to improved learning for many students. If you will simply read, sign, and date the attached form and have your child return it as soon as possible to his/her teacher, you will be helping us to begin our efforts.

Thank you for your attention. Sincerely,

Dr. Roger Azevedo, University of Maryland, College of Education, Dept. of Human Development 3304 Benjamin Building, Room 3304E, College Park, MD, 20742 Tel: 301-405-2799 E-mail: ra109@umail.umd.edu	Jennifer Cromley, Graduate Student, University of Maryland, College of Education, Dept. of Human Development 3304 Benjamin Building, College Park, MD, 20742 Tel: 301-405-2820 E-mail: jcromley@aol.com
---	--



PARENT PERMISSION FORM

Partic. ID _____

I consent to my child's participation in research project entitled:

Reading Processes of High School Students

Dr. Roger Azevedo, Principal Investigator, has explained the purpose of the study, the procedures to be followed and the extent of my child's participation. I understand that participation in this study is fully voluntary and that I may withdraw my child from the study at any time.

I acknowledge that my child's participation will occur in accordance with the established policies and procedures of the University of Maryland, [name of county] County's Public School System, and [name of high school]. I understand that my child may receive community service hours for participating in the research project. The information obtained about my child will remain anonymous and will be used only for research purposes.

I acknowledge that all information collected in the study is confidential, and my child's name will not be identified at any time. A numeric code will be used as identification on all data collection materials (i.e., audio recordings and standardized reading measures). I understand that the audio recordings and measures will allow Dr. Azevedo to investigate the reading processes used by high school students. Once data are collected, this code will be used for maintenance and analysis of data. Pseudonyms will be used in publications and conference papers.

There are no known risks to my child from this study. I understand that the study is not designed to benefit my child, but to help researchers understand high school students' reading processes, which may lead to developing better instructional methods, which can benefit other students in the future.

Finally, I acknowledge that I have read and fully understood this consent form. I sign it freely and voluntarily. A copy will be placed in my child's file.

Date: _____

Printed Name of Authorized Person: _____
(Person Authorized to Consent for Student)

Relation to Child: : _____

Signed: _____
(Person Authorized to Consent for Student)

Signed: _____
(Researcher)

Name of Student: _____

* Optional —Please indicate whether you are interested in having your child participate in future research related to this project on learning about high school students' reading by providing your name and phone number.

Name: _____ Phone Number: _____

Contact Information: Jennifer Cromley, Graduate Student, Tel: 301-405-2820 E-mail: jcromley@aol.com
Dr. Roger Azevedo, Assistant Professor, Tel: 301-405-2799 E-mail: ra109@umail.umd.edu
University of Maryland, College of Education, Dept. of Human Development, 3304 Benjamin Building, room 3304E, College Park, MD, 20742,

Contact Information: University of Maryland's Institutional Review Board, Lee Building #2133, Tel: 301-405-4212.

Attachment B— Student Assent Form

INFORMED ASSENT FORM Partic. ID _____

Identification of Project:	Reading Processes of High School Students
Statement of Age:	I state that I am under 18 years of age, in good physical health, and wish to participate in a program of research being conducted by Roger Azevedo and Jennifer Cromley of the University of Maryland, College Park, and in which my parents have already given permission for me to participate.
Purpose:	The purpose of this study is to understand high school students' reading processes, including strategy use, word reading, reading comprehension, vocabulary, and background knowledge.
Procedures:	I will complete standardized word reading, reading comprehension, vocabulary, reading strategy, and background knowledge measures. In addition, I may be selected to "think aloud" while reading a short passage from a high school social studies textbook and recall the passage, all of which will be audiotaped. The first session will last approximately 90 minutes and the second session about 45 minutes. The researchers will also collect any 8th or 9th grade reading test scores in my file.
Confidentiality:	All information collected in the study is confidential and my name will not be identified at any time.
Risks:	I understand that there are no risks associated with this assessment.
Benefits, Freedom to Withdraw, & Ability to Ask Questions:	I understand that the experiment is not designed to help me personally. I understand that I am free to ask questions at any time and withdraw from participation without penalty.
Name, Address, and Phone Number of Faculty Advisor:	Dr. Roger Azevedo, Assistant Professor, University of Maryland, College of Education, Dept. of Human Development, Benjamin Building, Room 3304E, College Park, MD, 20742 Tel: 301-405-2799E-mail: ra109@umail.umd.edu
The purpose and procedures of this study have been described to me and I give my assent to participate:	Name: _____ Date: _____ Signature: _____ (optional)



Student Background Information Sheet

Participant ID: _____

Date: _____

Teacher: _____

Name: _____

Class period: _____

Zip code where you live: _____

Room: _____

What grade are you in? _____

Have you ever been in Special Education? Yes No

Have you ever been held back a grade in school? Yes No

Parents' occupation:

Mother _____

Father _____

Highest level of parents' education:

Mother (circle one): Finished high school College Graduate school (Master's Degree or PhD)

Did not finish high school

Father (circle one): Finished high school College Graduate school (Master's Degree or PhD)

Did not finish high school

What language(s) did you speak growing up? _____

How old are you? _____

Race (please circle one or more) Asian Black Hispanic/Latino

Indian/Native Middle Eastern White/Caucasian Other _____

Are you (circle one): Female Male

When did you last study the Revolutionary War in school? _____ grade _____ month

APPENDIX D

Preliminary Study Background Knowledge Measure (from Cromley & Azevedo, 2004a)

Background knowledge measure

Please choose the one best answer for each question and fill in the corresponding circle on the answer sheet. PLEASE DO NOT WRITE ON THE TEST.

Sample: USA stands for:

- A United we Stand, America
- B Union Station Architecture
- C United States of America
- D Ukelele Symphony Anthem

1. A four-leaf clover is believed to
 - A Attract deer
 - B Be a sign of evil
 - C Be good to eat
 - D Bring good luck
2. A yam is
 - E A vegetable
 - F A furry animal
 - G A type of house
 - H A kind of pot
3. A caterpillar is
 - I A bird
 - J A kind of truck
 - K A tree
 - L An insect
4. A stove pipe is
 - M A kind of bird
 - N A kind of chimney
 - O A pipe for smoking tobacco
 - P A way to light a fire
5. A valve is
 - Q A kind of tool
 - R Carpenter's hardware
 - S Something that opens and closes
 - T Something valuable
6. Customs, duties or tariffs means
 - A A kind of law
 - B A tax on imported goods
 - C A right that a person has
 - D Serving in the military or other national service
7. The American colonies were:
 - E Other countries occupied by the United States
 - F The original states in the US
 - G Settlements in America controlled by Great Britain
 - H The Northern or Union states
8. Parliament is
 - I The king and his advisors
 - J A brand of cigarettes
 - K The House of Lords
 - L Like Congress, but in England
9. The Stamp Act was:
 - M A kind of stamp
 - N An American custom
 - O Action taken by the post office
 - P A law
10. Birch means
 - Q A type of tree
 - R A kind of stick
 - S An animal
 - T Large
11. Great Britain is in
 - A Europe
 - B North America
 - C Scandinavia
 - D The Soviet Bloc
12. Clover grows in
 - E Fresh water
 - F Greenhouses
 - G Lawns
 - H The ocean

13. Ice is
 I Always made by nature
 J Always made by people
 K Colorful
 L Frozen water
14. Jump rope is
 M A kind of twine
 N A playground or neighborhood game
 O An Olympic sport
 P Part of physical therapy
15. Baby turtles
 Q Grow in a pouch
 R Are born live from their mother
 S Hatch from eggs
 T Are born in winter
16. Repeal means
 A To fire a politician
 B To cancel a law
 C To tax
 D To pass a law
17. The Declaration of Independence was signed in
 E 1860
 F 1789
 G 1776
 H 1667
18. "No taxation without representation" means
 I Taxes are too high; working people cannot afford to pay them and still support a family
 J Taxes are wrong and they should never be allowed
 K It is not fair to have a representative in Congress unless you pay taxes
 L It is not fair to tax people unless they can elect someone to the legislature that passes the tax laws
19. The Revolutionary War was between:
 M The French and Indians
 N The North and South
 O The American colonies and Britain
 P Britain and Spain
20. In Medieval times there was/were no
 Q Cars or trucks
 R Cities or towns
 S Government
 T Books
21. In the old days (before 1800) there was/were no
 A Matches
 B Plumbing
 C Roads
 D Stores
22. A democracy is
 E A capitalist country
 F A country without political repression
 G A country where there are elections and people have rights
 H Only found in the United States
23. Before the Declaration of Independence, laws in America were made by
 I The President
 J Parliament
 K The Continental Congress
 L Washington, DC
24. After it rains over land, **most** of the water
 M Is soaked up by the land
 N Goes into water pipes
 O Runs off the land into lakes and streams
 P M and O
25. Medieval means the times from about
 Q 0 – 400 A.D.
 R 400 – 1400 A.D.
 S 1600 – 1800 A.D.
 T 1800 – 1900 A.D.
26. The top of a plant tub is
 A Round
 B Sharp
 C Sticky
 D Wet
27. In order for eggs to hatch,
 E People must break the shell
 F The baby animal must break the shell
 G The mother must crack the shell open
 H The shell breaks by itself
28. Wood is easy to light on fire if it is
 I Dark-colored
 J Hot
 K Dry
 L Old

29. In the Declaration of Independence,
- M The United States declared its independence from England
 - N The settlers declared their independence from the Indians
 - O The colonies declared their independence from Britain
 - P The Pilgrims declared their independence from France

30. A patriot is
- Q A good person
 - R A movie
 - S A person over 18
 - T Someone who is loyal to their country

31. Neutral means:
- A Not getting involved
 - B On the right side
 - C On the wrong side
 - D White

32. Shakespeare was
- E A Medieval scholar
 - F A modern novelist
 - G A scientist
 - H An Elizabethan playwright

33. Turtles make nests
- I Buried under dirt or sand
 - J High in trees or bushes
 - K Out of sticks and other plant material
 - L Underwater

APPENDIX E

Preliminary Study Inference Measure (from Cromley & Azevedo, 2004a)

Inference measure

Please read the passage and choose the one best answer for each question and fill in the corresponding circle on the answer sheet. PLEASE DO NOT WRITE ON THE TEST

So far as I know, [picking a four-leaf clover] was her only superstition, or anyway, the only one she ever acted on. And it was always used for the same purpose, which was to get my father's patients to pay their bills.

Very few of the patients paid promptly, and a good many never paid at all. Some sent in small checks, once every few months. A few remarkable and probably well-off patients paid immediately, the whole bill at once, and when this happened my father came upstairs after office hours greatly cheered.

- | | |
|---|---|
| <p>1. In the second sentence, what does "it" refer to?</p> <ul style="list-style-type: none">a. the cloverb. his father's billc. picking a four-leaf cloverd. his father's patient | <p>2. In the last sentence, why was his father "greatly cheered"?</p> <ul style="list-style-type: none">a. because someone paid their bill on timeb. because his mother found a four-leaf cloverc. because someone paid in cashd. because someone finished paying on layaway |
|---|---|
-

Sugar, Stamp, and Quartering Acts

Since the new Sugar Act would not afford a large revenue, it was supplemented in 1765 by the Stamp Act. This measure levied a direct tax on all newspapers printed in the colonies. It also taxed most commercial and legal documents used in business. It was realized that these two revenue acts would provide less than half the money needed for the army.

- | | |
|--|---|
| <p>3. Which of the following documents did <i>not</i> require a tax stamp?</p> <ul style="list-style-type: none">a. a newspaperb. a deedc. a bill of saled. a personal letter | <p>4. The last sentence suggests that the income from the acts would affect England's ability to</p> <ul style="list-style-type: none">a. protect the coloniesb. provide social servicesc. support merchantsd. build new roads |
|--|---|
-

That night [Alexandra] had a new consciousness of the country, felt almost a new relation to it. Even her talk with the boys had not taken away the feeling that had overwhelmed her when she drove back to the Divide that afternoon. She had never known before how much the country meant to her. The chirping of the insects down in the long grass had been like the sweetest music. She had felt as if her heart were hiding down there, somewhere, with the quail and the plover and all the little wild things that crooned or buzzed in the sun. Under the long shaggy ridges, she felt the future stirring.

5. What does the author mean by the phrase “as if her heart were hiding down there”?

- a. Alexandra was scared of the boys
- b. She did not want to feel overwhelmed and divided
- c. She did not like the noise of the insects
- d. She felt a strong connection to the land and its animals

6. The last sentence suggests that Alexandra was feeling

- a. Optimistic about what was to come
- b. Afraid of what might happen
- c. Calm and peaceful
- d. Scared because she was “under the long shaggy ridges”

A Stamp Act Congress, representing nine colonies, met in New York City on Oct. 7, 1765. The congress declared that only the colonial assemblies should tax the colonists. The congress also petitioned the king and Parliament for repeal of the objectionable measures. When the stamped papers began to arrive, mobs seized them or forced the ships' captains to take them back to England.

7. The colonial assemblies refers to a governing body in

- a. England
- b. Canada
- c. the American colonies
- d. the French colonies

8. The last sentence suggests that the mob was

- a. happy to see the ships arrive
 - b. delighted to see their family members
 - c. anxiously waiting for their letters
 - d. protesting the Stamp Act
-

The bottom egg had hatched with the others, but this female snapper had had a longer journey out of the nest. Battling upward through cast-off shells, she was tired. Now the crow saw her, climbing out of the sand, blinking in sunlight. He hopped up, wheeled and dived.

9. Why was the turtle tired?
- a. Because she had just hatched
 - b. She had to climb from the bottom of a deep nest
 - c. She knew there was a crow watching her
 - d. Because it took her longer to hatch out of her egg
10. The phrase “blinking in sunlight” suggests that
- a. The snapper was blind
 - b. The crow was blinking
 - c. The bottom of the nest was in bright sunlight
 - d. The bottom of the nest was in shade

Many wealthy merchants favored stopping all business that required the use of stamped papers. This, they said, would be perfectly legal. They also argued that it would so seriously interfere with the business of British merchants that Parliament would be forced to repeal the law.

11. What does “this” in the second sentence refer to?
- a. stamped papers
 - b. stopping all business
 - c. stopping all business that required stamped papers
 - d. wealthy merchants
12. Merchants thought Parliament would be forced to repeal the law because
- a. British business was suffering
 - b. wealthy merchants were displeased
 - c. all business had stopped
 - d. people were protesting

“Don’t feel bad,” said Yolonda, suddenly generous.
“Turning the ropes correctly is an art—it’s really hard.”
The bell rang and they both turned hurriedly toward the school.
“You have to have good rhythm and your partner has to be in sync with you. You know, really good vibes,” hollered Yolonda after Shirley’s scurrying figure. Without looking around, the Shirley person flapped her hand in a wave.
Well, I’ve impressed one person in this burg at least, thought Yolonda.

13. The author suggests that, in order to do Double Dutch well, partners have to

- a. Work well together
- b. Be strong
- c. Be fast
- d. Impress each other

14. Why did Yolonda think she had impressed Shirley?

- a. Because Shirley did not turn around when she waved
- b. Because Yolonda knew a lot about turning ropes and shared it with Shirley
- c. Because she bought Shirley a burger
- d. Because Yolonda hollered

The Outcry Against the Stamp Act

Opposition to the Stamp Act spread through the colonial assemblies. It came to a head in the Stamp Act Congress of 1765. The congress asserted that the colonists, as English subjects, could not be taxed without their consent. Alarmed by the refusal of the colonial towns to buy additional goods while the act remained in force, British merchants petitioned Parliament for its repeal.

15. Why did the Stamp Act Congress assert that the colonists could not be taxed?

- a. because they were good people
- b. because they were English subjects
- c. because they had not consented
- d. because they opposed the Stamp Act

16. What does “its” at the end of the last sentence refer to?

- a. Parliament
- b. the Stamp Act
- c. the Sugar Act
- d. British merchants

But when springtime comes and leaves begin to sprout, the [caterpillar] eggs hatch. Each little caterpillar cuts a round opening in the top of the egg with its jaws. Crawling out, it huddles on the twig with its newly hatched brothers and sisters. Soon the caterpillars spin a little silk tent for a community shelter. They leave it only at mealtime. Whenever it leaves the nest, each caterpillar spins a silk thread as a glistening trail behind it. Finished eating, the caterpillar follows the trail of silk.

17. Why does the caterpillar cut a hole in the egg?

- a. Because it is hungry and the egg is nourishing
- b. Because it wants to crawl
- c. To make silk out of it
- d. To get out of the egg when it is ready to hatch

18. Where does the caterpillar follow the trail of silk?

- a. Back home to the shelter
- b. To the leaves
- c. To its mother
- d. Back into its egg

I took a bite, finding it as sweet and hot as any I'd ever had, and was overcome with such a surge of homesickness that I turned away to keep my control. I walked along, munching the yam, just as suddenly overcome by an intense feeling of freedom—simply because I was eating while walking along the street. It was exhilarating. I no longer had to worry about who saw me or about what was proper. To hell with all that, and as sweet as the yam actually was, it became like nectar with the thought. If only someone who had known me at school or at home would come along and see me now.

19. What does the author take a bite of?

- a. Corn on the cob
- b. Fried chicken
- c. A yam
- d. It's impossible to tell

20. In the past, the author worried about someone seeing him

- a. eating in public
- b. eating a yam
- c. drinking nectar
- d. walking down the street

APPENDIX F

Preliminary Study Think-Aloud Practice Text (from Roller, 1986)

Practice text

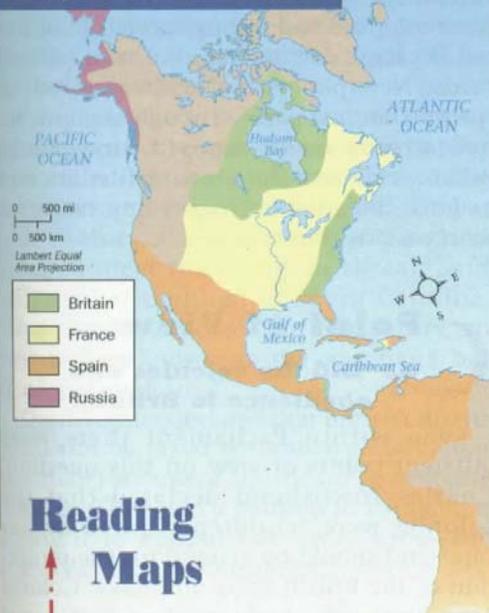
About two-thirds of the people of Indonesia live on the island of Java, the political heart of the country. ■ Java is about as large as the state of New York, but sixty million people live there. ■ New York has less than twenty million, yet we think of it as a very populous state. ■ If the Javanese were spread out over their island, there would be more than one thousand on each square mile. ■ But by no means is every square mile of Java habitable. ■ There are many mountains, including more than one hundred volcanoes, seventeen of which are active. ■

From Roller (1986)

APPENDIX G

Think-Aloud Protocol Text (from Viola, Wheatley & Hart, 1998)

European Claims 1750



European Claims 1763



Reading Maps

1. What parts of North America did the four European powers claim in 1750?
2. Describe the large transfer of territory that took place as a result of the British victory in the French and Indian War.

Without their usual help from the French, the tribes could not fight for long. They finally signed peace treaties with the British. The Indians had made Amherst pay dearly, though. They had captured or killed nearly 2,000 Englishmen.

The Proclamation of 1763 Parliament had to face the question of how to prevent further conflict with the Indians. It came up with an answer in the Proclamation of 1763. This **proclamation**—which means an official announcement—said that colonists could not settle west of the Appalachians. Settlers already living there were ordered to leave. In addition, the British government would take control of the fur trade.

The law was meant to buy time until Britain could sign treaties with the Indians. Land buyers, fur traders, and settlers, however, saw things differently. Now that the French were defeated and Pontiac's Rebellion had ended, they were eager to push westward. Most ignored the proclamation.

Parliament Taxes the Colonies

To enforce the Proclamation of 1763 and protect the colonists from the Indians, Britain declared that it needed an army of 10,000 soldiers in North America. But who should pay to support them? Taxpayers in Britain were already burdened with the debt

from the French and Indian War. The new British prime minister, George Grenville, looked for a way to make the colonies pay more for their own defense.

The Sugar Act Grenville knew that Parliament had never directly taxed the colonies. The Navigation Acts had only regulated trade so the colonies would do most of their buying and selling with Britain. Colonial merchants were to pay **customs duties**—charges on foreign imports—in order to sell non-British goods. However, merchants usually avoided the duties by bribing officials or smuggling.

One product often smuggled from the French West Indies was molasses, which was used for making rum, especially in New England. Grenville decided that enforcing the duty on foreign molasses would be a good way to raise **revenue**—income. In 1764 Parliament passed the Sugar Act, which cut in half the duty on foreign molasses to encourage merchants to pay it. However, the law also gave officials new powers to crack down on smuggling.

Under the Sugar Act, customs officials could enter any building at any time, using general search warrants called **writs of assistance**. Colonists accused of smuggling would face a panel of British judges instead of a jury of their peers. Also, they would be considered guilty unless proven innocent. Many colonists saw these searches and the denial of a jury trial as threats to their rights as Englishmen.

Meanwhile, Parliament's effort to get revenue went against the colonists' belief that they could only be taxed by their own legislatures. Protests against "taxation without representation," however, came mainly from New England merchants, who were most affected by the Sugar Act.

The Stamp Act To get even more revenue to pay for the colonies' defense,

Parliament passed the Stamp Act in 1765. Now colonists had to buy special stamps to put on legal documents, dice, and playing cards. Newspapers had to be printed on special stamped paper. Although stamp taxes had already been imposed for years in Britain, this was Parliament's first attempt to force the colonies to pay any tax other than customs duties.

Point of View

Did the colonies owe obedience to Britain?

Even within Parliament there were different points of view on this question. Charles Townshend declared that the colonies were "children planted by our care" and should be grateful for the protection of the British army and navy. Colonel Isaac Barré, who had served under General Wolfe in the French and Indian War, angrily responded,

"They planted by your care? No! Your oppressions planted them in America. They fled from your tyranny to a then uncultivated and inhospitable country. . . And believe me, remember I this day told you so, that same spirit of freedom . . . will accompany them still."

Britain would soon feel the effects of that spirit of freedom in the colonies.

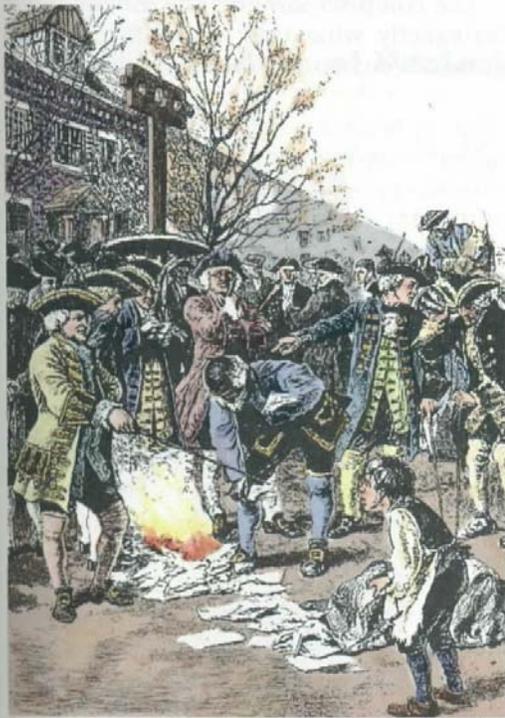
Protesting the Stamp Act

Unlike the Sugar Act, the Stamp Act affected people of every colony and social class, including leaders such as lawyers, newspaper publishers, and ministers. Protestors argued that only representatives they elected should be able to tax them. Such men lived near them and understood their

needs, unlike Parliament, which met thousands of miles across the sea. The protestors insisted that Parliament was taking their money against their will.

Concern was so widespread that nine colonies sent delegates to a meeting in New York called the Stamp Act Congress. The delegates saw the need for the colonies to put aside rivalries over land claims and trade in order to meet the common threat. Said Christopher Gadsden of South Carolina, "There ought to be no New Englanders, no New Yorkers known on the continent, but all of us Americans."

The delegates declared that British citizens could not be taxed without representation and that Parliament did not represent the colonies. They sent a petition to Parliament asking it to **repeal**—do away with—the Stamp Act and the Sugar Act.



Many colonists took matters into their own hands. Calling themselves Sons and Daughters of Liberty, they decided to **boycott**—refuse to buy—British goods. Daughters of Liberty agreed to wear homespun wool dresses rather than buy imported British cloth. Such pledges, known as non-importation agreements, were common throughout the colonies.

Some protests turned ugly. Tax collectors complained about "Sons of Violence" who tried to pressure them by breaking windows in their houses and even threatening their lives. Boston tax collector Andrew Oliver got a grim warning when an effigy—a dummy—of him was hanged and burned. A crowd in Connecticut even started to bury a tax collector alive. Only after hearing dirt being shoveled onto the coffin lid did he agree to resign.

Britain tried to tax the colonists by requiring them to buy special stamps for all legal documents and for newspapers. Tax collectors were appointed to sell the stamps, but most colonists refused to buy them. In one form of protest, they burned stamped papers.



APPENDIX H

Classes, Descriptions and Examples of the Variables Used to Code Learners' Components and Comprehension (based on Azevedo, Cromley, & Seibert, 2004).

Variable	Description	Example
Background knowledge		
Accurate PKA+	Activates accurate prior knowledge	“molasses is like syrup I think.”
Inaccurate PKA-	Activates inaccurate prior knowledge	“the African-American bus boycott when they refused to buy- to go on the buses to hurt the economy.”
ANACH-	Anachronism; believes that something from the present (e.g., a telephone) existed in the 1760’s	“people in like, like a conference room trying to get the Stamp Act and Sugar Act away.”
Inferences		
Accurate EVAL+	Any moral judgment about what is happening in the text, or an inference about moral judgments of people/groups in the passage	“they’re getting way too out of hand” “they weren’t given a fair trial” “I agree with that.” “I think that’s very awkward, because . . . that’s wrong!”
HYP+	Any hypothesis or prediction about events to follow in the text (could be an accurate, sensible, or even insensible hypothesis/prediction)	“ <u>But who should pay to support them?</u> I’m thinking us, not Britain.”
INF+	Makes an accurate within-text inference	“They’re trying to keep their land safe from the Indians I guess.” “So the British couldn’t pay- couldn’t really pay . . . because of the French and Indian War.”
KE+	Makes an accurate inference from PK + text	“the British needed big- a huge army.”

Variable	Description	Example
LINK+	Makes an accurate connection between what is read and an event from (recent or far) past	“I am thinking that this reminds me of . . . what we’re learning in history now, which was, or a while ago, which was when they would break into buildings when they thought that someone was communist.”
Inaccurate INF-	Makes an inaccurate within-text inference	“ <u>Britain would soon feel the effects of that spirit of freedom in the colonies.</u> So I guess he did a lot.”
KE-	Makes an inaccurate inference from PK + text	“So the Stamp Act affected more people like lawyers and newspaper publishers because they’re the ones who need to, uh, mail more stuff and send papers”
LINK-	Makes an inaccurate connection between what is read and an event from (recent or far) past.	“Don’t they do that in all kinds of courts? That you’re innocent until proven guilty.”
Strategies Accurate BTF+	“Back to the Future”—participant imagines him/herself in the past, and states what he or she would do in that situation.	“He held out longer than I would have.” “If I was a woman then, I wouldn’t want to wear no dress”
FOK+	States that he/she does understand, or shows evidence of understanding that is not a SUM+, INF+, KE+, etc.	“That makes sense.” “ <u>an effigy—a dummy—of him was hanged and burned.</u> . . . Uh, that- that’s a threat, I guess. . . Yeah. That is a threat.”
IMAGE+	States an accurate mental image of the situation (does not need to be relevant, just not inaccurate)	“I’m thinking of like soldiers walking.”

Variable	Description	Example
RR+	Re-reads 3 or more words in a row	<u>“To enforce the Proclamation of 1763 . . . I have to read it over again. To enforce the Proclamation of 1763”</u>
SQ+	Participant poses him/herself a question that might potentially be answered by the text; not a JOL; not PKA-	“Now I’m wondering what do their dresses look like?”
SUM+	Accurately summarizes (note that part of a summary can be accurate and part can be inaccurate—2 codes)	“So George Grenville, the British Prime Minister, was looking for other ways for them to pay for their own defense.”
Inaccurate JOL-	States that he/she does not understand	“I’m kind of confused.” <u>“But who should pay to support them? I don’t know!”</u>
IMAGE-	States an inaccurate mental image of the situation (that is not an anachronism)	“Thinking of jury and Englishmen talking, you know, trying”
NOTHINK-	Says “Nothing” or “I’m not thinking,” but not “no comment”	“I’m not really thinking nothing.” “nothing really, just what I read”
SUM-	Inaccurately summarizes (note that part of a summary can be inaccurate and part can be accurate—2 codes). Could misunderstand or over-generalize in the summary.	“So the Stamp Act had been around for a while but they weren’t really enforced”

Variable	Description	Example
Vocabulary		
Accurate VOC+	Shows directly (e.g., by rephrasing a definition) that the meaning of a word was understood—only code the first instance for each word	“I guess customs duties were the, uh, were like taxes.”
Inaccurate VOC-	Shows or states that the meaning of a word was unknown or misunderstood—only code the first instance for each word	“Rivalries, what are those?” “So the colonists would have to face judges instead of like people they know, or their friends.” “What’s molasses?”
Word reading		
Inaccurate WORD-	Mispronounces/miscues/substitutes/inserts a word in a way that affects meaning (e.g., “revenue” for “revenue”)	“Col-o-nel- col-o-nel <u>merchants</u> ” “ <u>needs, like</u> Par-lee-ament”
OMIT-	Omits a word(s) in a way that affects meaning (e.g., “they did [not] want to pay”)—omissions indicated in transcript in brackets	omits: [<u>a dummy—of him was hanged and burned</u>].

APPENDIX I

*Dissertation Study Parent Cover Letter, Parental Consent Form, Student Assent Form,
and Student Background Information Sheet*

Attachment A—Parental permission form

LETTER TO PARENTS

September 1, 2003

Dear Parent/Guardian:

With your permission, we would like to include your child in a study of reading processes used by high school students. The study, which will be conducted at [name of high school], has been approved by University of Maryland's Institutional Review Board (IRB) and your child's teacher, [name of teacher]. We expect the study to be completed by the end of June 2004.

The purpose of this study is to examine high school students' reading processes using standardized measures and by asking some children to "think aloud" while reading. Very little is known about reading processes in high school students, and we hope to learn more from students at various levels of reading proficiency. Your child's teacher may approve this activity for community service credit.

The main method we will use in the study is to complete brief word reading, fluency, reading comprehension, reading strategy, background knowledge, inference and vocabulary measures. In addition, one out of every six students will be selected to "think aloud" while reading a short passage from a high school social studies textbook and recalling the passage, all of which we will audiotape. There will be one session of approximately 90 minutes for all students and a second session of 20 minutes for the one-sixth of the students who participate in the "think-aloud" portion of the study. The study will be conducted at the school, either during class time or after school. We will then conduct a detailed analysis of the measures and audiotapes.

We will take a number of steps to ensure your child's privacy and anonymity. For example, no individuals' names will be used in any written or oral presentations resulting from the study and only a limited number of professional researchers will have access to the original tapes and in-class tests. All audiotapes will be destroyed after 5 years. Each student's participation will be fully voluntary. We will describe the general purpose of the study to each student so that each student can decide whether to participate and s/he may withdraw from the study at any time.

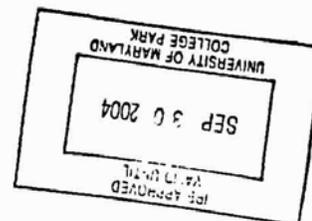
We hope that you will consent to your child's participation in this study that may one day contribute to improved learning for many students. If you will simply read, sign, and date the attached form and have your child return it as soon as possible to his/her teacher, you will be helping us to begin our efforts.

Thank you for your attention.

Sincerely,

Dr. Roger Azevedo, University of Maryland,
College of Education, Dept. of Human
Development
3304 Benjamin Building, Room 3304E,
College Park, MD, 20742
Tel: 301-405-2799
E-mail: ra109@cumail.umd.edu

Jennifer Cromley, Doctoral Candidate,
University of Maryland,
College of Education, Dept. of Human
Development
3304 Benjamin Building,
College Park, MD, 20742
Tel: 301-314-2670
E-mail: jcromley@aol.com



PARENT PERMISSION FORM

Partic. ID _____

Consent for Participation in Educational Research

I consent to my child's participation in research project entitled:

Reading Comprehension Component Processes in Early Adolescence

Dr. Roger Azevedo, Principal Investigator, has explained the purpose of the study, the procedures to be followed and the extent of my child's participation. I understand that participation in this study is fully voluntary and that I may withdraw my child from the study at any time.

I acknowledge that my child's participation will occur in accordance with the established policies and procedures of the University of Maryland, [name of county] County's Public School System, and [name of high school]. I understand that my child may receive community service hours for participating in the research project. The information obtained about my child will remain anonymous and will be used only for research purposes.

I acknowledge that all information collected in the study is confidential, and my child's name will not be identified at any time. A numeric code will be used as identification on all data collection materials (i.e., audio recordings and standardized reading measures). I understand that the audio recordings and measures will allow Dr. Azevedo to investigate the reading processes used by high school students. Once data are collected, this code will be used for maintenance and analysis of data. Pseudonyms will be used in publications and conference papers, and all audiotapes will be destroyed after 5 years.

There are no known risks to my child from this study. I understand that the study is not designed to benefit my child, but to help researchers understand high school students' reading processes, which may lead to developing better instructional methods, which can benefit other students in the future.

Finally, I acknowledge that I have read and fully understood this consent form. I sign it freely and voluntarily. A copy will be placed in my child's file.

Date: _____

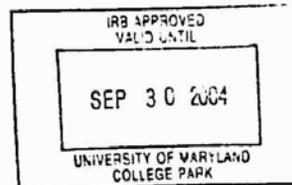
Printed Name of Authorized Person: _____
(Person Authorized to Consent for Student)

Relation to Child: _____

Signed: _____
(Person Authorized to Consent for Student)

Signed: _____
(Researcher)

Name of Student: _____



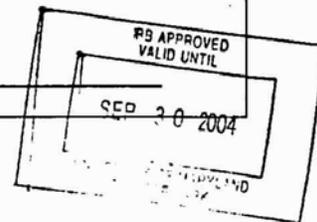
Contact Information: Jennifer Cromley, Doctoral Candidate, Tel: 301-314-2670 E-mail: jercromley@aol.com
Dr. Roger Azevedo, Assistant Professor, Tel: 301-405-2799 E-mail: ra109@umail.umd.edu
University of Maryland, College of Education, Dept. of Human Development, 3304 Benjamin Building, Room 3304E, College Park, MD, 20742
Contact Information: University of Maryland's Institutional Review Board, Lee Building #2133, Tel: 301-405-4212.

Attachment B—Student Assent form

INFORMED ASSENT FORM

Partic. ID _____

Identification of Project:	Reading Comprehension Component Processes in Early Adolescence
Statement of Age:	I state that I am under 18 years of age, in good physical health, and wish to participate in a program of research being conducted by Dr. Roger Azevedo and Jennifer Cromley of the University of Maryland, College Park, and in which my parents have already given permission for me to participate.
Purpose:	The purpose of this study is to understand high school students' reading processes, including strategy use, word reading, fluency, inference, reading comprehension, vocabulary, and background knowledge.
Procedures:	I will complete standardized word reading, fluency, reading comprehension, vocabulary, inference, reading strategy, and background knowledge measures. In addition, I may be selected to "think aloud" while reading a short passage from a high school social studies textbook and recall the passage, all of which will be audiotaped. The first session will last approximately 90 minutes and the second session approximately 20 minutes.
Confidentiality:	All information collected in the study is confidential and my name will not be identified at any time.
Risks:	I understand that there are no risks associated with this assessment.
Benefits, Freedom to Withdraw, & Ability to Ask Questions:	I understand that the experiment is not designed to help me personally. I understand that I am free to ask questions at any time and withdraw from participation without penalty.
Name, Address, and Phone Number of Faculty Advisor:	Dr. Roger Azevedo, Assistant Professor, University of Maryland, College of Education, Dept. of Human Development, Benjamin Building, Room 3304E, College Park, MD, 20742 Tel: 301-405-2799 E-mail: ra109@umail.umd.edu
The purpose and procedures of this study have been described to me and I give my assent to participate:	Name: _____ Date: _____ Signature: _____ (optional)



Student Background Information Sheet

Participant ID: _____

School: _____

Date: _____

Teacher: _____

Name: _____

Class period: _____

How old are you? _____

Room: _____

Race (please circle one or more) Asian Black Hispanic/Latino(a)
Indian/Native Middle Eastern White/Caucasian Other _____

Are you (circle one): Female Male

APPENDIX J

Passage used for One-Minute Fluency Measure (from Leslie & Caldwell, 2000)

World War I—Part I

World War I, also known as the Great War, drew in not only the major powers of Europe, but those of America and Asia as well. Many economic and political factors caused the war. Newly industrialized nations competed with one another for trade and markets for their goods. Also, the urge for national power and independence from other nations came from old and new powers. When a new nation tried to increase its power by building a strong military, an older nation perceived the new nation as a threat to its power. Such tensions led to the division of Europe into two groups for security: one composed of Britain, France, and Russia, the other of Austria, Hungary, and Germany.

Although the factors discussed above caused the war, the final breaking point was a local conflict between Austria and Serbia, a tiny kingdom in southeastern Europe. Serbia, supported by Russia, wanted to unite with the Serbs living in the Austro-Hungarian Empire and create a Greater Serbia. Austria, supported by Germany, did not want Serbia cutting into its empire. The war officially started in August of 1914, after the assassination of the Austrian heir to the throne, who was visiting Sarajevo, near Serbia's border. The assassin was a young man with connections to the military intelligence branch of the Serbian government. Austria's attempt to punish Serbia drew Russia and its allies Britain and France into a war against Austria—Hungary and Germany.

APPENDIX K

Background Knowledge Test (adapted from Cromley & Azevedo, 2004a)

Background knowledge measure

Please choose the one best answer for each question and fill in the corresponding circle on the answer sheet. PLEASE DO NOT WRITE ON THE TEST.

Sample: USA stands for:

- A United we Stand, America
 - B Union Station Architecture
 - C United States of America
 - D Ukelele Symphony Anthem
-

1. A stove pipe is
 - M A kind of bird
 - N A kind of chimney
 - O A pipe for smoking tobacco
 - P A way to light a fire
2. A valve is
 - Q A kind of tool
 - R Carpenter's hardware
 - S Something that opens and closes
 - T Something valuable
3. The American colonies were:
 - E Other countries occupied by the United States
 - F The original states in the US
 - G Settlements in America controlled by Great Britain
 - H The Northern or Union states
4. The Stamp Act was:
 - M A kind of stamp
 - N An American custom
 - O Action taken by the post office
 - P A law
5. Birch means
 - Q A type of tree
 - R A kind of stick
 - S An animal
 - T Large
6. Clover grows in
 - E Fresh water
 - F Greenhouses
 - G Lawns
 - H The ocean
7. Jump rope is
 - M A kind of twine
 - N A playground or neighborhood game
 - O An Olympic sport
 - P Part of physical therapy
8. The Declaration of Independence was signed in
 - E 1860
 - F 1789
 - G 1776
 - H 1667
9. "No taxation without representation" means
 - I Taxes are too high; working people cannot afford to pay them and still support a family
 - J Taxes are wrong and they should never be allowed
 - K It is not fair to have a representative in Congress unless you pay taxes
 - L It is not fair to tax people unless they can elect someone to the legislature that passes the tax laws
10. The Revolutionary War was between:
 - M The French and Indians
 - N The North and South
 - O The American colonies and Britain
 - P Britain and Spain
11. In Medieval times there was/were no
 - Q Cars or trucks
 - R Cities or towns
 - S Government
 - T Books
12. Before the Declaration of Independence, laws in America were made by
 - I The President
 - J Parliament
 - K The Continental Congress
 - L Washington, DC

13. After it rains over land, **most** of the water
- M Is soaked up by the land
 - N Goes into water pipes
 - O Runs off the land into lakes and streams
 - P M and O
14. Medieval means the times from about
- Q 0 – 400 A.D.
 - R 400 – 1400 A.D.
 - S 1600 – 1800 A.D.
 - T 1800 – 1900 A.D.
15. The top of a plant tub is
- A Round
 - B Sharp
 - C Sticky
 - D Wet
16. In order for eggs to hatch,
- E People must break the shell
 - F The baby animal must break the shell
 - G The mother must crack the shell open
 - H The shell breaks by itself
17. Wood is easy to light on fire if it is
- I Dark-colored
 - J Hot
 - K Dry
 - L Old
18. In the Declaration of Independence,
- M The United States declared its independence from England
 - N The settlers declared their independence from the Indians
 - O The colonies declared their independence from Britain
 - P The Pilgrims declared their independence from France
19. Shakespeare was
- E A Medieval scholar
 - F A modern novelist
 - G A scientist
 - H An Elizabethan playwright
20. Turtles make nests
- I Buried under dirt or sand
 - J High in trees or bushes
 - K Out of sticks and other plant material
 - L Underwater

APPENDIX L

Selected Items from the Vocabulary Measure

Selected Items from the Vocabulary Measure

Note: Questions 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, and 34 are the odd-numbered questions from the Gates-MacGinitie Vocabulary subtest (Fourth Edition, Level 7/9 Form S), and in order to protect copyright are *not* reproduced here.

Practice Question

V-1. a big garage

- K place for cars
- L machine
- M sidewalk
- N covered porch
- O cloth sack

-
3. the officials
- K. employers
 - L. government representatives
 - M. not fakers
 - N. rude people
 - O. spokesmen

15. they represent
- K. groom
 - L. look nice
 - M. pay for
 - N. speak for
 - O. talk about

6. many legislatures
- F. governments
 - G. governors
 - H. law-making bodies
 - I. mansions
 - J. people

18. the Parliament
- F. king and his advisors
 - G. a business group
 - H. cigarettes
 - I. House of Lords
 - J. like Congress, but in England

9. he must enforce
- A. go in
 - B. beat up
 - C. crack down on
 - D. finish up
 - E. force into

21. the customs duties
- A. business laws
 - B. taxes on imports
 - C. personal rights
 - D. national service
 - E. strong fabric

12. she had a petition
- P. letter
 - Q. meeting
 - R. protest
 - S. put-down
 - T. request

24. he was burdened with
- P. bothered by
 - Q. concerned by
 - R. loaded down with
 - S. tired of
 - T. worried about

27. far-away colonies

- K. farms
- L. houses
- M. states
- N. territories
- O. villages

33. They got some molasses

- A. chemical
- B. fabric
- C. oil
- D. sweet syrup
- E. vegetable

30. they repealed

- F. fired
- G. requested
- H. canceled
- I. taxed
- J. passed

APPENDIX M

Inference and Strategy Use Test (adapted from Cromley & Azevedo, 2004a)

Inference and Strategy Use measure

Please read the passage and choose the one best answer for each question and fill in the corresponding circle on the answer sheet. PLEASE DO NOT WRITE ON THE TEST

So far as I know, picking a four-leaf clover was her only superstition, or anyway, the only one she ever acted on. And it was always used for the same purpose, which was to get my father's patients to pay their bills.

Very few of the patients paid promptly, and a good many never paid at all. Some sent in small checks, once every few months. A few remarkable and probably well-off patients paid immediately, the whole bill at once, and when this happened my father came upstairs after office hours greatly cheered.

1. In the second sentence, what does "it" refer to?
 - A. the clover
 - B. his father's bill
 - C. picking a four-leaf clover
 - D. his father's patient
 2. In the last sentence, why was his father "greatly cheered"?
 - E. because someone paid their bill on time
 - F. because his mother found a four-leaf clover
 - G. because someone paid in cash
 - H. because someone finished paying on layaway
 3. Which of the following is most likely to follow this passage?
 - I. An explanation of why the author's mother picked a four-leaf clover
 - J. How much the average doctor bill was at the time
 - K. What the family spent the money on when patients paid on time
 - L. Why the author's father was happy
 4. Which of the following would be most useful to know in order to understand the passage?
 - M. The author is writing about the Great Depression
 - N. "Her" refers to the author's mother
 - O. Rich people pay their bills on time
 - P. Doctors are happy when patients pay their bills
-

CONTINUED ON NEXT PAGE

Sugar, Stamp, and Quartering Acts

Since the new Sugar Act would not afford a large revenue, it was supplemented in 1765 by the Stamp Act. This measure levied a direct tax on all newspapers printed in the colonies. It also taxed most commercial and legal documents used in business. It was realized that these two revenue acts would provide less than half the money needed for the army.

5. Which of the following documents did **NOT** require a tax stamp?
- A. a newspaper
 - B. a deed
 - C. a bill of sale
 - D. a personal letter
6. The last sentence suggests that the income from the acts would affect England's ability to
- E. protect the colonies
 - F. provide social services
 - G. support merchants
 - H. build new roads
7. Which of the following is the best summary of the paragraph?
- I. The Sugar Act did not raise much money
 - J. The Stamp Act taxed newspapers and legal documents
 - K. Armies cost money to support
 - L. The Stamp Act was passed before 1776
8. What would be the best strategy for finding the date the Stamp Act was passed?
- M. Re-read the entire passage
 - N. Skim the paragraph for dates
 - O. Ask a friend
 - P. Use the Table of Contents

CONTINUED ON NEXT PAGE

That night Alexandra had a new consciousness of the country, felt almost a new relation to it. Even her talk with the boys had not taken away the feeling that had overwhelmed her when she drove back to the Divide that afternoon. She had never known before how much the country meant to her. The chirping of the insects down in the long grass had been like the sweetest music. She had felt as if her heart were hiding down there, somewhere, with the quail and the plover and all the little wild things that crooned or buzzed in the sun. Under the long shaggy ridges, she felt the future stirring.

9. What does the author mean by the phrase “as if her heart were hiding down there”?
- A. Alexandra was scared of the boys
 - B. She did not want to feel overwhelmed and divided
 - C. She did not like the noise of the insects
 - D. She felt a strong connection to the land and its animals
10. The last sentence suggests that Alexandra was feeling
- E. Optimistic about what was to come
 - F. Afraid of what might happen
 - G. Calm and peaceful
 - H. Scared because she was “under the long shaggy ridges”
11. Which of the following questions could **NOT** be answered from the passage?
- I. What kind of animals live in the country?
 - J. What was the weather like that day?
 - K. How did Alexandra feel?
 - L. What kind of music did Alexandra like?
12. Which of the following sentences could most easily be omitted from the paragraph without changing its meaning?
- M. That night Alexandra had a new consciousness of the country. . .
 - N. Even her talk with the boys had not taken away the feeling . . .
 - O. She had never known before how much the country meant . . .
 - P. The chirping of the insects down in the long grass had been . . .

CONTINUED ON NEXT PAGE

A Stamp Act Congress, representing nine colonies, met in New York City on Oct. 7, 1765. The congress declared that only the colonial assemblies should tax the colonists. The congress also petitioned the king and Parliament for repeal of the objectionable measures. When the stamped papers began to arrive, mobs seized them or forced the ships' captains to take them back to England.

13. The colonial assemblies refers to a governing body in
- A. England
 - B. Canada
 - C. the American colonies
 - D. the French colonies
14. The last sentence suggests that the mob was
- E. happy to see the ships arrive
 - F. delighted to see their family members
 - G. anxiously waiting for their letters
 - H. protesting the Stamp Act
15. Which of the following is the best summary of the passage?
- I. The Stamp Act Congress met in 1765
 - J. Mobs seized stamped papers from English ships
 - K. The Stamp Act Congress met and asked Parliament to repeal the Act
 - L. Nine colonies met and petitioned the King
16. From the context, the word "objectionable" is probably
- M. a positive word
 - N. a negative word
 - O. a neutral word
 - P. irrelevant to the passage

CONTINUED ON NEXT PAGE

The bottom egg had hatched with the others, but this female snapper had had a longer journey out of the nest. Battling upward through cast-off shells, she was tired. Now the crow saw her, climbing out of the sand, blinking in sunlight. He hopped up, wheeled and dived.

17. Why was the turtle tired?
- A. Because she had just hatched
 - B. She had to climb from the bottom of a deep nest
 - C. She knew there was a crow watching her
 - D. Because it took her longer to hatch out of her egg
18. The phrase “blinking in sunlight” suggests that
- E. The snapper was blind
 - F. The crow was blinking
 - G. The bottom of the nest was in bright sunlight
 - H. The bottom of the nest was in shade
19. Which of the following questions could **NOT** be answered from the passage?
- I. Was the snapper turtle male or female?
 - J. What color was the crow?
 - K. Had the turtle’s eyes opened?
 - L. Could the crow fly?
20. Which of the following is most likely to follow this passage?
- M. The turtle’s hatching
 - N. A description of the crow’s nest
 - O. The crow trying to catch the turtle
 - P. The turtle climbing out of the nest

CONTINUED ON NEXT PAGE

Many wealthy merchants favored stopping all business that required the use of stamped papers. This, they said, would be perfectly legal. They also argued that it would so seriously interfere with the business of British merchants that Parliament would be forced to repeal the law.

21. What does “this” in the second sentence refer to?
- A. stamped papers
 - B. stopping all business
 - C. stopping all business that required stamped papers
 - D. wealthy merchants
22. Merchants thought Parliament would be forced to repeal the law because
- E. British business was suffering
 - F. wealthy merchants were displeased
 - G. all business had stopped
 - H. people were protesting
23. Which sentence would it make the most sense to underline or highlight?
- I. Many wealthy merchants favored stopping all business . . .
 - J. This, they said, would be . . .
 - K. They also argued that it would so seriously interfere . . .
 - L. All of the above.
24. Another way to express the last sentence is:
- M. Parliament was forced to repeal the law because it interfered with British merchants’ business.
 - N. They argued that it would interfere with the merchants of British business and Parliament would be forced to repeal the law.
 - O. The boycott would interfere with the business of British merchants.
 - P. They argued that British merchants’ business would be disrupted so much that Parliament would be forced to repeal the law.

CONTINUED ON NEXT PAGE

“Don’t feel bad,” said Yolonda, suddenly generous.
“Turning the ropes correctly is an art—it’s really hard.”
The bell rang and they both turned hurriedly toward the school.
“You have to have good rhythm and your partner has to be in sync with you. You know, really good vibes,” hollered Yolonda after Shirley’s scurrying figure. Without looking around, the Shirley person flapped her hand in a wave.
Well, I’ve impressed one person in this burg at least, thought Yolonda.

25. The author suggests that, in order to do Double Dutch well, partners have to

- A. Work well together
- B. Be strong
- C. Be fast
- D. Impress each other

27. Which of the following is most likely to follow this passage?

- I. Yolonda walks back to school by herself
- J. Yolonda walks home by herself
- K. Yolonda gives Shirley money
- L. Shirley turns around

26. Why did Yolonda think she had impressed Shirley?

- E. Because Shirley did not turn around when she waved
- F. Because Yolonda knew a lot about turning ropes and shared it with Shirley
- G. Because she bought Shirley a burger
- H. Because Yolonda hollered

28. Which of the following would be most useful to know in order to understand the passage?

- M. Turning ropes means Double Dutch or jump rope
- N. People usually turn around when they wave
- O. Waving to someone is a sign of friendliness
- P. Vibes and sync are terms from jazz music

CONTINUED ON NEXT PAGE

I took a bite, finding it as sweet and hot as any I'd ever had, and was overcome with such a surge of homesickness that I turned away to keep my control. I walked along, munching the yam, just as suddenly overcome by an intense feeling of freedom—simply because I was eating while walking along the street. It was exhilarating. I no longer had to worry about who saw me or about what was proper. To hell with all that, and as sweet as the yam actually was, it became like nectar with the thought. If only someone who had known me at school or at home would come along and see me now.

29. What does the author take a bite of?

- A. Corn on the cob
- B. Fried chicken
- C. A yam
- D. It's impossible to tell

30. In the past, the author worried about someone seeing him

- E. eating in public
- F. eating a yam
- G. drinking nectar
- H. walking down the street

31. Which of the following is the best summary of the passage?

- I. The author had mixed feelings after eating a yam
- J. The author felt homesick while eating familiar food
- K. The ate a yam and drank some nectar while walking down the street
- L. The author felt free doing something he was never allowed to do growing up

32. Which sentence best captures the main idea of the passage?

- M. I took a bite, finding it as sweet and hot as any I'd ever had . . .
- N. I walked along, munching the yam, just as suddenly overcome . . .
- O. I no longer had to worry about who saw me or about what
- P. To hell with all that, and as sweet as the yam actually was . . .

APPENDIX N

Think-Aloud Practice Text (adapted from Roller, 1986)

Practice text

About two-thirds of the people of Indonesia live on the island of Java, the political heart of the country. ■ Java is about as large as the state of New York, but 123 million people live there. ■ New York has less than twenty million, yet we think of it as a very populous state. ■ If the Javanese were spread out over their island, there would be more than two thousand on each square mile. ■ But by no means is every square mile of Java habitable. ■ There are many mountains, including more than one hundred volcanoes, seventeen of which are active. ■

Adapted from Roller (1986)

APPENDIX O

Think-Aloud Codes, Definitions, and Examples (adapted from Azevedo, Cromley & Seibert, 2004; Azevedo, Guthrie & Seibert, 2004; Cromley & Azevedo, 2004b)

<i>Code</i> <i>[Major coding category]</i>	<i>Definition</i>	<i>Example</i>
ANACH- [BKGD-]	Anachronism; believes that something from the present or more recent past existed in the 1760's.	"People started to like threaten the Congress."
BTF+ [INF+]	"Back to the Future"—participant imagines him/herself in the past, and states what he or she would do in that situation.	"I would have been concerned, too."
COIS+ [STRAT+]	Coordinates text and picture.	" <u>In one form of protest, they burned stamped papers.</u> [Looks at picture] They burned the stamped paper."
EVAL+ [INF+]	Any moral judgment about what is happening in the text, or an inference about moral judgments of people/groups in the passage.	"How nice!"
FOK+ [STRAT+]	States that he/she does understand, or shows evidence of understanding that is not a SUM+, INF+, KE+, VOC+, etc. Do not code "OK," "All right" as FOK+	"OK, that makes sense."
HYP+ [INF+]	Any hypothesis or prediction about events to follow in the text (could be an accurate, sensible, or even insensible hypothesis/prediction).	"They might start, like, charging people for the soldiers."
IMAGE+ [STRAT+]	States an accurate mental image of the situation (does not need to be relevant, just not inaccurate).	"I'm think that, like British judges look mean."
INF+ [INF+]	Makes an accurate within-text inference. The word "because" always signals an inference, but not all inferences include a causal word.	"Because they had a lot of goods to sell."
INF- [INF-]	Makes an inaccurate within-text inference.	"So they were doing illegal acts in order to raise money."

<i>Code</i> [Major coding category]	<i>Definition</i>	<i>Example</i>
JOL- [STRAT-]	States that he/she does not understand.	“I don’t understand none of this!”
KE+ [INF+]	Makes an accurate inference by putting together prior knowledge and text.	“ <u>But who should pay to support them?</u> Like who is going to buy all of the weapons.”
KE- [INF-]	Makes an inaccurate inference by putting together prior knowledge and text.	“ <u>But who should pay to support them?</u> Other people in the towns.”
LINK+ [INF+]	Makes an accurate connection between what is read and an event from (recent or far) past.	“So they thought of the writs of assistance and all that as . . . a threat to their rights, civil rights.”
LINK- [INF-]	Makes an inaccurate connection between what is read and an event from (recent or far) past.	“Don’t they do that in all kinds of courts? That you’re innocent until proven guilty.”
NOTHINK- [STRAT-]	Says “Nothing” or “I’m not thinking,” but not “no comment” or “I don’t know.”	“Well, I’m not thinking of anything right now.”
OMIT- [WORD-]	Omits a word in a way that affects meaning—omissions indicated in transcript in brackets.	“ <u>A [crowd in Connecticut even started to bury a tax collector] alive.</u> ”
PKA+ [BKGD+]	Activates prior knowledge that is both accurate and relevant.	“But didn’t Britain like force them to take it? Yeah.”
PKA- [BKGD-]	Activates inaccurate and/or irrelevant prior knowledge or states that he/she lacks background knowledge.	“That rum was like prohibited, like in those times.”
RR+ [STRAT+]	Re-reads 5 or more words in a row.	“ <u>Britain declared that it needed any—an army of 10,000 soldiers in North America. Britain declared that it needed an army of 10,000 soldiers in North America.</u> ”

<i>Code</i> [Major coding category]	<i>Definition</i>	<i>Example</i>
SC- [WORD-]	Miscues on a word, then self-corrects (at any later re-reading of the same word). Includes people's names.	" <u>and protect the colonies—colonists.</u> "
SQ+ [STRAT+]	Participant poses him/herself a question that might potentially be answered by the forthcoming text; not a rhetorical question; need not include a question word; not a misunderstanding of text that came before (e.g., INF-, VOC-); not JOL-; not PKA-.	"I'm thinking how big is the army then if they needed 10,000 soldiers just for North America."
SUM+ [STRAT+]	Accurately summarizes (note that part of a summary can be accurate and part can be inaccurate—2 codes).	"to make people pay for their own defenses."
SUM- [STRAT-]	Inaccurately summarizes (note that part of a summary can be inaccurate and part can be accurate—2 codes). Could misunderstand or over-generalize in the summary; not PKA-.	"They were selling their things."
TN+ [STRAT+]	Takes notes.	"To enforce . the Proclamation writing it all down."
VOC+ [VOC+]	Shows directly (e.g., by rephrasing a definition) that the meaning of a word was understood—only code the first instance for each word.	"So alcohol, rum."
VOC- [VOC-]	Shows or states that the meaning of a word was unknown or misunderstood—only code the first instance for each word.	" <u>In order to meet the common threat . . .</u> so they tried to meet in the middle somewhere."
WORD- [WORD-]	Mispronounces/miscues/substitutes/ inserts a word in a way that affects meaning. Count only the first miscue for any particular word, even if later read incorrectly in a different way. No people's names (e.g., Greenville for Grenville) are counted as WORD-.	"The nivation Acts."

<i>Code</i> <i>[Major coding category]</i>	<i>Definition</i>	<i>Example</i>
Do not code	“OK,” “Oh,” “all right,” “no comment,” “uh,” “yeah,” “...,” “I guess,” “I think,” “I don’t know,” miscues or omissions that do not affect meaning, interest statements.	

Note: Underlining indicates student is reading from the text.

APPENDIX P

Verbal recall a priori Coding Rubric

Verbal Recall a priori Coding Rubric

Major topics (4 points each)	Sub-topics (2 points each)	Supporting details (1 point each)
To fund army		French and Indian War British debt need money protect colonists 10,000 soldiers Indians
Sugar Act	halved/dropped [tax] duties/tax [on molasses] molasses enforce	1760s rum West Indies writs of assistance non-British goods smuggling bribing British judges guilty until proven innocent
No taxation without representation	no representatives	Parliament 1,000s of miles away
Stamp Act	buy tax stamps [not just special stamp] on paper affected more people disagreements in Parliament	tax stamp legal documents dice playing cards newspapers Barre Townshend children fled planted

Major topics (4 points each)	Sub-topics (2 points each)	Supporting details (1 point each)
Stamp Act (cont'd.)	Stamp Act Congress	9 colonies New York unite petition repeal
	Protests/dislike/rebel	violent windows Sons of Violence threaten/own hands effigy Oliver burn [effigy or paper] hang [effigy] [CT] tax collector started to bury resigned boycotts women cloth English goods made own Sons and Daughters of Liberty

REFERENCES CITED

- Aarnoutse, C., & van Leeuwe, J. (2000). Development of poor and better readers during the elementary school. *Educational Research and Evaluation, 6*(3), 251-278.
- Adams, B. C., Bell, L. C., & Perfetti, C. A. (1995). A trading relationship between reading skill and domain knowledge in children's text comprehension. *Discourse Processes, 20*(3), 307-323.
- Afflerbach, P. (1990). The influence of prior knowledge and text genre on readers' prediction strategies. *Journal of Reading Behavior, 22*(2), 131-148.
- Afflerbach, P. (2000). Verbal reports and protocol analysis. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3) (pp. 163-179). Mahwah, NJ: Erlbaum.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*(3), 317-332.
- Alexander, P. A., & Jetton, T. L. (2000). Learning from text: A multidimensional and developmental perspective. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3) (pp. 285-310). Mahwah, NJ: Erlbaum.
- Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research, 58*(4), 375-404.

- Alexander, P. A., & Kulikowich, J. M. (1991). Domain knowledge and analogic reasoning ability as predictors of expository text comprehension. *Journal of Reading Behavior, 23*(2), 165-190.
- Alexander, P. A., Murphy, P. K., Woods, B. S., Duhon, K. E., & Parker, D. (1997). College instruction and concomitant changes in students' knowledge, interest, and strategy use: A study of domain learning. *Contemporary Educational Psychology, 22*(2), 125-146.
- Alibali, M. W., Bassok, M., Solomon, K. O., Syc, S. E., & Goldin-Meadow, S. (1999). Illuminating mental representations through speech and gesture. *Psychological Science, 10*(4), 327-333.
- Allington, R. (December, 2002). *What really matters for struggling readers*. Paper presented at the Research-Based Best Practices in Literacy Instruction conference, Miami, FL.
- Alvermann, D. E., Hinchman, K. A., Moore, D. W., Phelps, S. F., & Waff, D. R. (Eds.) (1998). *Reconceptualizing the literacies in adolescents' lives*. Mahwah, NJ: Erlbaum.
- Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. (1977). Frameworks for comprehending discourse. *American Educational Research Journal, 14*, 362-382.
- Artelt, C., Schiefele, U., & Schneider, S. (2001). Predictors of reading literacy. *European Journal of Psychology of Education, 16*(3), 262-383.
- Azevedo, R., & Cromley, J.G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*(3), 523-535.

- Azevedo, R., Cromley, J. G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology, 29*, 344-370.
- Azevedo, R., Guthrie, J.T., & Seibert, D. (2004). Do different instructional interventions facilitate students' ability to shift to more sophisticated mental models of complex systems? *Journal of Educational Computing Research, 30*(1), 87-111.
- Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluid reading. *Journal of Memory and Language, 24*, 119-131.
- Baker, L., & Cerro, L. C. (2000). Metacognitive skills and reading. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 99-146). Lincoln, NE: Buros Institute.
- Battin-Pearson, S., Newcomb, M. D., Abbott, R. D., Hill, K. G., Catalano, R. F., & Hawkins, J. D. (2000). Predictors of early high school dropout: A test of five theories. *Journal of Educational Psychology, 92*(3), 568-582.
- Baumann, J. F. (1984). The effectiveness of a direct instruction paradigm for teaching main idea comprehension. *Reading Research Quarterly, 20*(1), 93-115.
- Baumann, J. F., Edwards, E. C., Boland, E. M., Olejnik, S., & Kame'enui, E. J. (2003). Vocabulary tricks: Effects of instruction in morphology and context on fifth-grade students' ability to derive and infer word meanings. *American Educational Research Journal, 40*(2), 447-494.
- Bean, T. W., Searles, D., Singer, H., & Cowen, S. (1990). Learning concepts from biology text through pictorial analogies and an analogical study guide. *Journal of Educational Research, 83*(4), 233-237.

- Beck, I., McKeown, M., Sandora, C., Kucan, L., & Worthy, J. (1996). Questioning the author: A year-long classroom implementation to engage students with text. *Elementary School Journal, 96*, 385-414.
- Beishuizen, J., Le Grand, J., & van der Schalk, J. (1999). No correlation between inferencing causal relations and text comprehension? *Learning and Instruction, 9*(1), 37-56.
- Bell, L. C., & Perfetti, C. A. (1994). Reading skill: Some adult comparisons. *Journal of Educational Psychology, 86*(2), 244-255.
- Bentler, P. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.
- Blachman, B. A. (2000). Phonological awareness. In M. L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3) (pp. 483-502). Mahwah, NJ: Erlbaum.
- Blanc, N., & Tapiero, I. (2001). Updating spatial situation models: Effects of prior knowledge and task. *Discourse Processes, 31*(3), 241-262.
- Bleeker, M. M., & Jacobs, J. E. (2004). Achievement in math and science: Do mothers' beliefs matter 12 years later? *Journal of Educational Psychology, 96*(1), 97-109.
- Bourassa, D. C., Levy, B. A., Dowin, S., & Casey, A. (1998). Transfer effects across contextual and linguistic boundaries: Evidence from poor readers. *Journal of Experimental Child Psychology, 71*(1), 45-61.
- Bracewell, R. J. & Bruleux, A. (1994). Substance and romance in analyzing think-aloud protocols. In P. Smagorinsky (Ed.), *Speaking about writing: Reflection on research methodology*. Newbury Park, CA: Sage.

- Bransford, J. D., & Johnson, M. (1973). Considerations of some problems of comprehension. In W.G. Chase (Ed.), *Visual Information Processing* (pp. 383-438). New York: Academic Press.
- Bråten, I., & Samuelstuen, M. S. (2004). Does the influence of reading purpose on reports of strategic text processing depend on students' topic knowledge? *Journal of Educational Psychology, 96*(2), 324-336.
- Breznitz, Z. (1997). Reading rate acceleration: Developmental aspects. *The Journal of Genetic Psychology, 158*(4), 427-441.
- Britton, B. K., & Gülgöz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology, 83*(3), 329-345.
- Britton, B. K., Stimson, M., Stennet, M., & Gülgöz, S. (1998). Learning from instructional text: Test of an individual-differences model. *Journal of Educational Psychology, 90*(3), 476-491.
- Brown, G. H., & Fetters, W. B. (1984). *Two years in high school: The status of 1980 sophomores in 1982*. Washington, DC: U.S. Department of Education.
- Brown, A. L., & Palincsar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 393-451). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bryant, D. P., Vaughn, S., Linan-Thompson, S., Ugel, N., Hamff, A., & Hougen, M. (2000). Reading outcomes for students with and without reading disabilities in general education middle-school content area classes. *Learning Disability Quarterly, 23*(4), 238-252.

- Buly, M. R., & Valencia, S. W. (2002). Below the bar: Students who fail state reading assessments. *Educational Evaluation and Policy Analysis, 24*(3), 219-239.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-281.
- Byrnes, J., & Guthrie, J. T. (1992). Prior conceptual knowledge and textbook search. *Contemporary Educational Psychology, 17*(1), 8-29.
- Caillies, S., Denhière, G., & Kintsch, W. (2002). The effect of prior knowledge on understanding from text: Evidence from primed recognition. *European Journal of Cognitive Psychology, 14*(2), 267-286.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*(1), 31-42.
- Callahan, D., & Drum, P. (1984). Reading ability and prior knowledge as predictors of eleven and twelve year olds' text comprehension. *Reading Psychology, 5*(1-2), 145-154.
- Carney, J. J., Anderson, D., Blackburn, C., & Blessing, D. (1984). Preteaching vocabulary and the comprehension of social studies materials by elementary school children. *Social Education, 48*(3), 195-196.
- Carnine, D. W., Kameenui, E. J., & Woolfson, N. (1982). Training of textual dimensions related to text-based inferences. *Journal of Reading Behavior, 14*(3), 335-340.
- Carr, S. C., & Thompson, B. (1996). The effects of prior knowledge and schema activation strategies on the inferential reading comprehension of children with and without learning disabilities. *Learning Disability Quarterly, 19*, 48-61.

- Carr, T. H., & Levy, B. A. (Eds.). (1990). *Reading and its Development*. San Diego: Academic Press.
- Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities, 36*(2), 151-164.
- Center on Education Policy. (2004). State high school exit exams: A maturing reform. Washington, DC: Author.
- Chan, L. K. S. (1994). Relationship of motivation, strategic learning, and reading achievement in grades 5, 7, and 9. *Journal of Experimental Education, 62*(4), 319-339.
- Chang, L., Liu, H., Wen, Z., Fung, K. Y., Wang, Y., & Xu, Y. (2004). Mediating teacher liking and moderating authoritative teaching on Chinese adolescents' perceptions of antisocial and prosocial behaviors. *Journal of Educational Psychology, 96*(2), 369-380
- Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- Chi, M. T. H., Hutchinson, J. E., & Robin, A. F. (1989). How inferences about novel domain-related concepts can be constrained by structured knowledge. *Merrill-Palmer Quarterly, 35*(1), 27-62.
- Christopherson, S. L., Schulz, C. B., & Waern, Y. (1981). The effect of two contextual conditions on recall of a reading passage and on thought processes in reading. *Journal of Reading, 23*, 573-578.

- Chung, G. K. W. K., de Vries, L. F., Cheak, A. M., Stevens, R. H., & Bewley, W. L. (2002). Cognitive process validation of an online problem solving assessment. *Computers in Human Behavior, 18*, 669-684.
- Clarke, M., Haney, W., & Madaus, G. (2000). High stakes testing and high school completion. *NBETPP Statements, 1*(3), 1-11.
- Cohen, J. (1983). Things I have learned (so far). *American Psychologist, 45*, 1304-1312.
- Coladarci, T., & McIntire, W. G. (1988). Gender, urbanicity, and ability. *Research in Rural Education, 5*(1), 13-15.
- Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1197-1211.
- Crain-Thoreson, C., Lippman, M. Z., & McClendon-Magnuson, D. (1997). Windows on comprehension: Reading comprehension processes as revealed by two think-aloud procedures. *Journal of Educational Psychology, 89*(4), 579-591.
- Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. N.Y.: Holt, Rinehart, and Winston.
- Cromley, J. G. & Azevedo, R. (2004a). *Testing the fit of three models of reading comprehension with 9th grade students*. Poster presented at the 2004 annual meeting of the American Educational Research Association, San Diego, CA.
- Cromley, J. G. & Azevedo, R. (2004b). *Using think-aloud data to illuminate a model of high school reading comprehension*. Poster presented at the 2004 annual meeting of the American Educational Research Association, San Diego, CA.

- Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology, 83*, 264-274.
- Cunningham, A. E., & Stanovich, K. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*(6), 934-945.
- Cunningham, A. E., Stanovich, K., & Wilson, M. R. (1990). Cognitive variation in adult college students differing in reading ability. In T. H. Carr & B. A. Levy (Eds.), *Reading and its Development* (pp. 129-160). San Diego: Academic Press.
- Curtis, M. E. (1980). Development of components of reading skill. *Journal of Educational Psychology, 72*(5), 656-669.
- Curtis, M. E. (2002). Adolescent reading: A synthesis of research. Manuscript submitted for publication to *Journal of Adolescent and Adult Literacy*.
- Dalaker, J. (2001). *Poverty in the United States: 2000*. Washington, DC: U.S. Bureau of the Census.
- Davey, B. (1987). Relations between word knowledge and comprehension: Generalization across tasks and readers. *Journal of Educational Research, 80*(3), 179-183.
- Davey, B. (1988). The nature of response errors for good and poor readers when permitted to reinspect text during question-answering. *American Educational Research Journal, 25*(3), 399-414.
- de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading, 6*(1), 51-77.

- De Soto, J. L., & De Soto, C. B. (1983). Relationship of reading achievement to verbal processing abilities. *Journal of Educational Psychology*, 75(1), 116-127.
- Dewitz, P., Carr, E. M., & Patberg, J. P. (1987). Effects of inference training on comprehension and comprehension monitoring. *Reading Research Quarterly*, 22(1), 99-121.
- Dixon, P., LeFevre, J.-A., & Twilley, L. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology*, 80(4), 465-472.
- Dole, J. A., Brown, K. J., & Trathen, W. (1996). The effects of strategy instruction on the comprehension performance of at-risk students. *Reading Research Quarterly*, 31(1), 62-88.
- Dole, J. A., Valencia, S. W., Greer, E. A., & Wardrop, J. L. (1991). Effects of two types of prereading instruction on the comprehension of narrative and expository text. *Reading Research Quarterly*, 26(2), 142-159.
- Dreher, M. J., & Brown, R. F. (1993). Planning prompts and indexed terms in textbook search tasks. *Journal of Educational Psychology*, 85(4), 662-669.
- Dreher, M. J., & Guthrie, J. T. (1990). Cognitive processes in textbook chapter search tasks. *Reading Research Quarterly*, 25(4), 323-339.
- Earthman, E. A. (1992). Creating the virtual work: Readers' processes in understanding literary texts. *Research in the Teaching of English*, 26(4), 351-384.
- Ehrlich, M.-F., Kurtz-Costes, B., & Loridant, C. (1993). Cognitive and motivational determinants of reading comprehension in good and poor readers. *Journal of Reading Behavior*, 25(4), 365-381.

- Eldredge, J. L., Quinn, B. & Butterfield, D. D. (1990). Causal relationships between phonics, reading comprehension, and vocabulary achievement in second grade. *Journal of Educational Research*, 83(4), 201-214.
- Ericsson, K. A. (Ed.). (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*. Mahwah, NJ: Erlbaum.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102(2), 211-245.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). Cambridge, MA: MIT Press.
- Ericsson, K. A., & Smith, J. (Eds.). (1991). *Toward a general theory of expertise: Prospects and limits*. New York: Cambridge University Press.
- Fehrenbach, C. (1991). Gifted/average readers: Do they use the same reading strategies? *Gifted Child Quarterly*, 35(3), 125-127.
- Feltovich, P. J., Ford, K. M., & Hoffman, R. H. (Eds.). (1997). *Expertise in context: Human and machine*. Cambridge, MA: MIT Press.
- Fincher-Kiefer, R. (1992). The role of prior knowledge in inferential processing. *Journal of Research in Reading*, 15(1), 12-27.
- Franks, B. A. (1997). Deductive reasoning with prose passages: Effects of age, inference form, prior knowledge, and reading skill. *International Journal of Behavioral Development*, 21(3), 501-535.
- Franks, B. A. (1998). Logical inference skills in adult reading comprehension: Effects of age and formal education. *Educational Gerontology*, 24, 47-68.

- Freebody, P., & Anderson, R. C. (1983a). Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. *Reading Research Quarterly*, 18(3), 277-294.
- Freebody, P., & Anderson, R. C. (1983b). Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Reading Behavior*, 15(3), 19-39.
- Garcia, G. E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *Reading Research Quarterly*, 25(4), 371-392.
- Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3), 430-445.
- Goldman, S. R., & Duran, R. P. (1988). Answering questions from oceanography texts: Learner, task, and text characteristics. *Discourse Processes*, 11, 373-412.
- 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: Processes and intervention* (pp. 1-13). Mahwah, NJ: Erlbaum.
- Graesser, A. C., & Bertus, E. L. (1998). The construction of causal inferences while reading expository texts on science and technology. *Scientific Studies of Reading*, 2(3), 247-269.
- Graesser, A. C., & Britton, B. K. (1996). Five metaphors for text understanding. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 341-351). Mahwah, NJ: Erlbaum.

- Graesser, A. C., Kassler, M. A., Kreuz, R. J., & McLain-Allen, B. (1998). Verification of statements about story worlds that deviate from normal conceptions of time: What is true about "Einstein's Dreams?" *Cognitive Psychology*, 35(3), 246-301.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48(1), 163-189.
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31(1), 104-137.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371-395.
- Graves, M. F., Boettcher, J. A., Peacock, J. L., & Ryder, R. J. (1980). Word frequency as a predictor of students' reading vocabularies. *Journal of Reading Behavior*, 12(2), 117-127.
- Graves, M. F., Cook, C. L., & LaBerge, M. J. (1983). Effects of previewing difficult short stories on low ability junior high school students' comprehension, recall, and attitudes. *Reading Research Quarterly*, 18(3), 262-276.
- Greene, J. F. (1998). Another chance: Help for older students with limited literacy. *American Educator*, 22(1-2), 74-79.
- Grigg, W. S. , Daane, M. C., Jin, Y., & Campbell, J. R. (2003). *The nation's report card: Reading 2002*. Washington, DC: U.S. Department of Education.
- Guthrie, J. T., Goldberg, H. K., & Finucci, J. (1972). Independence of abilities in disabled readers. *Journal of Reading Behavior*, 4(2), 129-138.
- Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research: Vol. III* (pp. 403-422). Mahwah, NJ: Erlbaum.

- Hadwin, A. F., & Winne, P. H. (2001). CoNoteS2: A software tool for promoting self-regulation. *Educational Research and Evaluation, 7*(2-3), 313-334.
- Hadwin, A., Winne, P., Stockley, D., Nesbit, J., & Woszczyzna, C. (2001). Context moderates students' self-reports about how they study. *Journal of Educational Psychology, 93*(3), 477-487.
- Haenggi, D., & Perfetti, C. A. (1994). Processing components of college-level reading comprehension. *Discourse Processes, 17*, 83-104.
- Haines, L. P., & Leong, C. K. (1983). Coding processes in skilled and less skilled readers. *Annals of Dyslexia, 33*, 67-89.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis (5th Ed.)*. Upper Saddle River, NJ: Prentice Hall.
- Haney, W., Madaus, G., Abrams, L., Wheelock, A., Miao, J., & Gruia, I. (2004). *The education pipeline in the United States, 1970-2000*. Boston, MA: The National Board on Educational Testing and Public Policy.
- Hannon, B., & Daneman, M. (1998). Facilitating knowledge-based inferences in less-skilled readers. *Contemporary Educational Psychology, 23*(2), 149-172.
- Hansen, J., & Pearson, P. D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. *Journal of Educational Psychology, 75*(6), 821-829.
- Hare, V. C., Rabinowitz, M., & Schieble, K. M. (1989). Text effects on main idea comprehension. *Reading Research Quarterly, 24*(1), 72-88.
- Harker, W. J. (1994). "Plain sense" and "poetic significance": Tenth-grade readers reading two poems. *Poetics, 22*, 199-218.

- Harris, T. L., & Hodges, R. E. (Eds.). (1995). *The literacy dictionary: The vocabulary of reading and writing*. Newark, DE: International Reading Association.
- Hayes, D. A., & Tierney, R. J. (1982). Developing readers' knowledge through analogy. *Reading Research Quarterly, 17*(2), 256-280.
- Hecht, S. A. (1999). Individual solution processes while solving addition and multiplication math facts in adults. *Memory & Cognition, 27* (6), 1097-1107.
- Hood, J., & Dubert, L. A. (1983). Decoding as a component of reading comprehension among secondary students. *Journal of Reading Behavior, 15*(4), 51-61.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.
- Huelsenbeck, J. H., Larget, B., & Alfaro, M. E. (2004). Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Molecular Biology and Evolution, 21*(6), 1123-1133.
- Ivey, G. (1999). A multicase study in the middle school: Complexities among young adolescent readers. *Reading Research Quarterly, 34*(2), 172-192.
- Jackson, M. D., & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General, 108*(2), 151-181.
- Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*(4), 719-729.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

- Kameenui, E. J., Carnine, D. W., & Freschi, R. (1982). Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. *Reading Research Quarterly, 17*(3), 367-388.
- Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction, 7*(3), 161-195.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*(4), 294-303.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.
- Kintsch, W., Britton, B. K., Fletcher, C. R., Kintsch, E., Mannes, S. M., & Nathan, M. J. (1993). A comprehension-based approach to learning and understanding. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation: Vol. 30* (pp. 165-214). New York: Academic Press.
- Kintsch, W., & van Dijk, T. A. (1978). Towards a model of text comprehension and production. *Psychological Review, 85*, 363-394.
- Klenk, L., & Kibby, M. W. (2000). Re-mediating reading difficulties: Appraising the past, reconciling the present, constructing the future. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research: Volume III* (pp. 667-690). Mahwah, NJ: Erlbaum.
- Kletzien, S. B. (1991). Strategy use by good and poor comprehenders reading expository text of differing levels. *Reading Research Quarterly, 26*(1), 67-86.

- Kletzien, S. B. (1992). Proficient and less-proficient comprehenders' strategy use for different top-level structures. *Journal of Reading Behavior, 24*(2), 191-215.
- Kline, R. (1998). *Principles and practice of structural equation modeling*. NY: Guilford.
- Konold, T. R., Juel, C., McKinnon, M., & Deffes, R. (2003). A multivariate model of early reading acquisition. *Applied Psycholinguistics, 24*, 89-112.
- Kozminsky, E., & Kozminsky, L. (2001). How do general knowledge and reading strategies ability relate to reading comprehension of high school students at different educational levels? *Journal of Research in Reading, 24*(2), 187-204.
- Kuhara-Kojima, K., & Hatano, G. (1991). Contribution of content knowledge and learning ability to the learning of facts. *Journal of Educational Psychology, 83*(2), 253-263.
- Kuhara-Kojima, K., Hatano, G., Saito, H., & Haebara, T. (1996). Vocalization latencies of skilled and less-skilled comprehenders for words written in hiragana and kanji. *Reading Research Quarterly, 31*(2), 158-171.
- Kunz, G. C., Drewniak, U., & Schott, F. (1992). On-line and off-line assessment of self-regulation in learning from instructional text. *Learning and Instruction, 2*, 287-301.
- Laing, S. P., & Kamhi, A. G. (2002). The use of think-aloud protocols to compare inferencing abilities in average and below-average readers. *Journal of Learning Disabilities, 35*(5), 436-447.
- Langer, J. A. (1980). Relation between levels of prior knowledge and the organization of recall. In M. L. Kamil & A. J. Moe (Eds.), *Perspectives in reading research and instruction* (pp. 28-33). Washington, DC: National Reading Conference.

- Langer, J. A. (1984). Examining background knowledge and text comprehension. *Reading Research Quarterly, 19*(4), 468-481.
- Langer, J. A. (1986). *Children reading and writing: Structures and strategies*. Norwood, NJ: Ablex.
- Lau, K., & Chan, D. W. (2003). Reading strategy use and motivation among Chinese good and poor readers in Hong Kong. *Journal of Research in Reading, 26*(2), 177-190.
- Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology, 95*(2), 211-224.
- Lee, C. (2001). Is October Brown Chinese? A cultural modeling activity system for underachieving students. *American Educational Research Journal, 38*(1), 97-141.
- Lee, C. D. (1995). A culturally based cognitive apprenticeship: Teaching African American high school students skills in literary interpretation. *Reading Research Quarterly, 30*(4), 608-630.
- Leslie, L., & Caldwell, J. (2000). *Qualitative reading inventory III*. New York: Longman.
- Leslie, L., & Cooper, J. (1993). Assessing the predictive validity of prior knowledge assessment. In D. J. Leu & C. K. Kinzer (Eds.), *Examining central issues in literacy research, theory and practice: Forty-second yearbook of the National Reading Conference* (pp. 93-100). Rochester, NY: National Reading Conference.
- Levesque, C., Zuehlke, A. N., Stanek, L. R., Ryan, R. M. (2004). Autonomy and competence in German and American university students: A comparative study based on self-determination theory. *Journal of Educational Psychology, 96*(1), 68-84.

- Levy, B. A., & Hinchley, J. (1990). Individual and developmental differences in the acquisition of reading skills. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development* (pp. 81-128). San Diego: Academic Press.
- Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis* (3rd ed.). Mahwah, NJ: Erlbaum.
- Long, D. L., & Bourg, T. (1996). Thinking aloud: Telling a story about a story. *Discourse Processes, 21*(3), 329-339.
- Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1456-1470.
- Long, D. L., Seely, M. R., & Oppy, B. J. (1999). The strategic nature of less skilled readers' suppression problems. *Discourse Processes, 27*(3), 281-302.
- Loxterman, J. A., Beck, I. L., & McKeown, M. G. (1994). The effects of thinking out loud during reading on students' comprehension of more or less coherent text. *Reading Research Quarterly, 29*(4), 353-367.
- Lundeberg, M.A. (1987). Metacognitive aspects of reading comprehension: Studying understanding in legal case analysis. *Reading Research Quarterly, 22*(4), 407-432.
- MacDonald, G W., & Cornwall, A. (1995). The relationship between phonological awareness and reading and spelling achievement eleven years later. *Journal of Learning Disabilities, 28*(8), 523-527.
- MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L.G. (2001). *Gates-MacGinitie reading tests, Level 7/9, Form S (4th Ed.)*. Itasca, IL: Riverside.

- MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L.G. (2002). *Technical report for the fourth edition, Gates-MacGinitie reading tests*. Itasca, IL: Riverside.
- Magliano, J. P., & Graesser, A. C. (1991). A three-pronged method for studying inference generation in literary text. *Poetics, 20*, 193-232.
- Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology, 91*(4), 615-629.
- Mahony, D. (1994). Using sensitivity to word structure to explain variance in high school and college level reading ability. *Reading and Writing: An Interdisciplinary Journal, 6*, 19-44.
- Mannes, S., & Kintsch, W. (1987). Knowledge organization and text organization. *Cognition and Instruction, 4*, 91-115.
- Margosein, C. M., Pascarella, E. T., & Pflaum, S. W. (1982). The effects of instruction using semantic mapping on vocabulary and comprehension. *Journal of Early Adolescence, 2*(2), 185-194.
- Marsh, H. W., Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96*(1), 56-67.
- Martino, N. L., & Hoffman, P. R. (2002). An investigation of reading and language abilities of college freshmen. *Journal of Research in Reading, 25*(3), 310-318.

- Mathewson, G. C. (1989). Effects of if-then usage upon urban students' inference generation during American history reading. In S. McCormick & J. Zutell (Eds.), *Cognitive and social perspectives for literacy research and instruction, 38th Yearbook of the National Reading Conference* (pp. 331-338). Chicago, IL: National Reading Conference.
- Mayer, R. E. (1998). *The promise of educational psychology: Learning in the content areas*. Upper Saddle River, NJ: Prentice Hall.
- McBride-Chang, C., Manis, F. R., Seidenberg, M. S., & Custodio, R. G. (1993). Print exposure as a predictor of word reading and reading comprehension in disabled and nondisabled readers. *Journal of Educational Psychology, 85*(2), 230-238.
- McKeown, M. G., Beck, I. L., Ohmanson, R. C., & Perfetti C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior, 15*(1), 3-18.
- McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology, 55*(1), 51-62.
- McNamara, D. S. (2003). SERT: Self-Explanation Reading Training. *Discourse Processes, 38*(1), 1-30.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1-43.
- Medo, M. A., & Ryder, R. J. (1993). The effects of vocabulary instruction on readers' ability to make causal connections. *Reading Research and Instruction, 33*(2), 119-134.

- Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980) Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly, 16*(1), 72-103.
- Miyake, N., & Norman, D. A. (1979). To ask a question, one must know enough to know what is not known. *Journal of Verbal Learning and Verbal Behavior, 18*(3), 357-364.
- Moats, L. M. (2000). *Speech to print: Language essentials for teachers*. Baltimore: Brookes Publishing.
- Moje, E. B. (2000). “*All the stories that we have*”: Adolescents’ insights about literacy and learning in secondary schools. Newark, DE: International Reading Association.
- Moje, E. B., & O’Brien, D. G. (Eds.) (2001). *Constructions of literacy: Studies of teaching and learning in and out of secondary schools*. Mahwah, NJ: Erlbaum.
- Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*(2), 249-259.
- Moore, D. W., Bean, T. W., Birdyshaw, D., & Rycik, J. A. (1999). *Adolescent literacy: A position statement*. Newark, DE: International Reading Association.
- Moore, P. J., & Scevak, J. J. (1997). Learning from texts and visual aids: A developmental perspective. *Journal of Research in Reading, 20*(3), 205-223.
- Mosborg, S. (2002). Speaking of history: How adolescents use their knowledge of history in reading the daily news. *Cognition and Instruction, 20*(3), 323-358.
- Murphy, S., Shannon, P., Johnston, P., & Hansen, J. (1998). *Fragile evidence: A critique of reading assessment*. Mahwah, NJ: Erlbaum.

- Muth, K. D., Glynn, B. K., Britton, B. K., & Graves, M. F. (1988). Thinking out loud while studying text: Rehearsing key ideas. *Journal of Educational Psychology*, 80(3), 315-318.
- Myers, J. L., & Well, A. D. (2003). *Research design and statistical analysis (2nd Ed.)*. Mahwah, NJ: Erlbaum.
- Nagy, W., Berninger, V., Abbott, R., Vaughan, K., & Vermuelen, K. (2003). Relationship of morphology and other language skills to literacy skills in at-risk second-grade readers and at-risk fourth-grade writers. *Journal of Educational Psychology*, 95(4), 730-742.
- Nagy, W. E., Diakidoy, I-A. N., & Anderson, R. C. (1993). The acquisition of morphology: Learning the contribution of suffixes to the meanings of derivatives. *Journal of Reading Behavior*, 25(2), 155-170.
- Nagy, W. E., & Scott, J. A. (2000). Vocabulary processes. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3) (pp. 483-502). Mahwah, NJ: Erlbaum.
- National Assessment Governing Board. (1999). *Reading framework for the National Assessment of Educational Progress, 1992-2000*. Washington, DC: Author.
- National Center for Education Statistics. (2000). *NAEP 1999 trends in academic progress: Three decades of student performance*. Washington, DC: Author.
- National Institute for Child Health and Human Development. (2002). *Adolescent literacy: Research needs*. Downloaded from the World Wide Web on November 19, 2002 from http://www.nifl.gov/partnershipforreading/adolescent/research_needs.html.

- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: NICHD.
- Neuman, S. B. (1990). Assessing children's inferencing strategies. In J. Zutell & J. McCormick (Eds.), *Literacy theory and research: Analyses from multiple paradigms* (pp. 267-274). Thirty-ninth yearbook of the National Reading Conference. Chicago, IL: National Reading Conference.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- No Child Left Behind Act of 2001. PL 107-110, 20 U.S.C. (2001).
- Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Quarterly Journal of Experimental Psychology: Learning, Memory, & Cognition*, 17(3), 398-415.
- Oakhill, J., & Yuill, N. (1996). Higher order factors in comprehension disability: Processes and remediation. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 69-92). Mahwah, NJ: Erlbaum.
- Olshavsky, J. E. (1976-77). Reading as problem solving: An investigation of strategies. *Reading Research Quarterly*, 12(4), 654-764.
- Olshavsky, J. E. (1978). Comprehension profiles of good and poor readers across materials of increasing difficulty. In P. D. Pearson (Ed.), *Reading: Disciplined inquiry in process and practice: 27th yearbook of the National Reading Conference* (pp. 73-76). Clemson, SC: National Reading Conference.

- Otero, J. T., & Campanario, J. M. (1990). Comprehension evaluation and regulation in learning from science texts. *Journal of Research in Science Teaching*, 27(5), 447-460.
- Otero, J. T., & Kintsch, W. (1992). Failures to detect contradictions in text: What readers believe versus what they read. *Psychological Science*, 3(4), 229-235.
- Owings, J. (1995). *Two years later: Cognitive gains and school transitions of NELS:88 eighth graders*. Washington, DC: U.S. Department of Education.
- Paris, S. G., & Lindauer, B. K. (1976). The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology*, 8, 217-227.
- Paris, S. R., & Oka, E. R. (1986). Children's reading strategies, metacognition, and motivation. *Developmental Review*, 6(1), 25-56.
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction (3rd Ed.)*. Fort Worth, TX: Harcourt Brace.
- Perfetti, C. (1985). *Reading ability*. NY: Oxford University Press.
- Perfetti, C. A. (1988). Verbal efficiency in reading ability. In M. Daneman, G. E. MacKinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice (Vol. 6)*, (pp. 109-143). NY: Academic Press.
- Perfetti, C. A. (1989). There are generalized abilities and one of them is reading. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 307-335). Hillsdale, NJ: Erlbaum.
- Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. Gough, L. C. Ehri, & R. Trieman (Eds.), *Reading Acquisition* (pp. 145-174). Hillsdale, NJ: Erlbaum.

- Perfetti, C. A. (1994). Psycholinguistics and reading ability. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 849-894). Dan Diego, CA: Academic Press.
- Perfetti, C. A., & Hart, L. (2001). The lexical bases of comprehension skill. In D. S. Gorfien (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67-86). Washington, DC: American Psychological Association.
- Perry, M., & Elder, A. D. (1997). Knowledge in transition: Adults' developing understanding of a principle of physical causality. *Cognitive Development, 12*, 131-157.
- Perry, N. (2002). Introduction: Using qualitative methods to enrich understandings of self-regulated learning. *Educational Psychologist, 37*(1), 1-4.
- Peskin, J. (1998). Constructing meaning when reading poetry: An expert-novice study. *Cognition and Instruction, 16*(3), 235-263.
- Peverly, S. T., Brobst, K. E., Graham, M., & Shaw, S. (2003). College adults are not good at self-regulation: A study on the relationship of self-regulation, note taking, and test taking. *Journal of Educational Psychology, 95*(2), 335-346.
- Phillips, L. M. (1988). Young readers' inference strategies in reading comprehension. *Cognition and Instruction, 5*(3), 193-222.
- Pintrich, P. (2000). The role of goal orientation and self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeichner (Eds.), *Handbook of self-regulation* (pp. 452-494). San Diego: Academic Press.
- Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3) (pp. 545-561). Mahwah, NJ: Erlbaum.

- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.
- Pritchard, R. (1990). The effects of cultural schemata on reading processing strategies. *Reading Research Quarterly*, 25(4), 273-295.
- Reardon, S., & Galindo, C. (April, 2002). *Do high-stakes tests affect students' decisions to drop out of school? Evidence from NELS*. Presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology*, 80(1), 16-20.
- Reynolds, R. E., Shephard, C., Lapan, R., Kreek, C., & Goetz, E. T. (1990). Differences in the use of selective attention by more successful and less successful tenth-grade readers. *Journal of Educational Psychology*, 82(4), 749-759.
- Roberts, T. A. (1988). Development of pre-instruction versus previous experience: Effects on factual and inferential comprehension. *Reading Psychology*, 9(2), 141-157.
- Robertson, W. C. (1990). Detection of cognitive structure with protocol data: Predicting performance on physics transfer problems. *Cognitive Science*, 14(2), 253-280.
- Rogers, T. (1991). Students as literary critics: The interpretive experiences, beliefs, and processes of ninth-grade students. *Journal of Reading Behavior*, 23(4), 391-423.
- Roller, C. M. (1986). Overcoming underwriting and other textbook sins. *Social Education*, 50(1), 56-57.
- Rosenshine, B., & Meister, C. (1997). Cognitive strategy instruction in reading. In S. A. Stahl & D. A. Hayes (Eds.), *Instructional models in reading* (pp. 85-107). Hillsdale, NJ: Erlbaum.

- Rouet, J.-F. (2003). What was I looking for? The influence of task specificity and prior knowledge on students' search strategies in hypertext. *Discourse Processes*, 3(2), 163-186
- Saarnio, D. A., Oka, E. R., & Paris, S. G. (1990). Developmental predictors of children's reading comprehension. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development* (pp. 57-80). San Diego: Academic Press.
- Sabatini, J. P. (2002). Efficiency in word reading of adults: Ability group comparisons. *Scientific Studies of Reading*, 6(3), 267-298.
- Saiegh-Haddad, E. (2003). Bilingual oral reading fluency and reading comprehension: The case of Arabic/Hebrew (L1)-English (L2) readers. *Reading and Writing: An Interdisciplinary Journal* 16, 717-736.
- Samuels, S. J. (1994). Toward a theory of automatic information processing in reading, revisited. In R. B. Rudell, M. R. Rudell, & H. Singer (Eds.) *Theoretical models and processes in reading* (4th Ed.), (pp. 816-837). Newark, DE: International Reading Association.
- Scarborough, H. S., Ehri, L. C., Olson, R. K., & Fowler, A. E. (1998). The fate of phonemic awareness beyond the elementary school years. *Scientific Studies of Reading*, 2, 115-142.
- Scardamalia, M., & Bereiter, C. (1992). Text-based and knowledge-based questioning by children. *Cognition & Instruction*, 9(3), 177-199.
- Schiefele, U. (1996). Topic interest, text representation, and quality of experience. *Contemporary Educational Psychology*, 21(3), 3-18.
- Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes*, 33(2), 105-132.

- Schmidt, H. G., De Volder, M. L., De Grave, W. S., Moust, J. H. C., & Patel, V. L. (1989). Explanatory models in the processing of science tests: The role of prior knowledge activation through small-group discussion. *Journal of Educational Psychology, 81*, 610-619.
- Schneider, W. (1993). Domain-specific knowledge and memory performance in children. *Educational Psychology Review, 5*(3), 257-273.
- Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology, 81*(3), 116-127.
- Schoonen, R., Hulstijn, J., & Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in grades 6, 8, and 10. *Language Learning, 48*(1), 71-106.
- Schulman, J. J., & Havighurst, R. J. (1947). Relationship between ability and social status in a midwestern community. IV: Size of vocabulary. *Journal of Educational Psychology, 38*, 437-442.
- Shankweiler, D., Lundquist, E., Dreyer, L. G., & Dickinson, C. C. (1996). Reading and spelling difficulties in high school students: Causes and consequences. *Reading and Writing: An Interdisciplinary Journal, 8*, 267-294.
- Sheehan, K. M., & Mislevy, R. J. (1990). Integrating cognitive and psychometric models in a measure of document literacy. *Journal of Educational Measurement, 27*, 255-272.

- Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology, 81*(1), 91-100.
- Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.
- Singer, M., & Halldorson, M. (1996). Constructing and validating motive bridging inferences. *Cognitive Psychology, 30*(1), 1-38.
- Singer, M., & Kintsch, W. (2001). Text retrieval: A theoretical exploration. *Discourse Processes, 31*(1), 27-59.
- Singer, M. H., & Crouse, J. (1981). The relationship of context-use skills to reading: A case for an alternative experimental logic. *Child Development, 52*(4), 1326-1329.
- Slater, W. H., Graves, M. F., & Piche, G. L. (1985). Effects of structural organizers on ninth-grade students' comprehension and recall of four patterns of expository text. *Reading Research Quarterly, 20*(2), 189-202.
- Smiley, S. S., Pasquale, F. L., & Chandler, C. L. (1976). The pronunciation of familiar, unfamiliar and synthetic words by good and poor adolescent readers. *Journal of Reading Behavior, 8*(3), 289-297.
- Smith, M. W., & Wilhelm, J. D. (2002). *"Reading don't fix no Chevys": Literacy in the lives of young men*. Portsmouth, NH: Heinemann.
- Snow, C. (Ed.). (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Snow, C.E. & Strucker, J. (2000). Lessons from *Preventing Reading Difficulties in Young Children* for adult learning and literacy. *Annual Review of Adult Learning and Literacy, 1*, 25-73.

- Spear-Swerling, L., & Sternberg, R. J. (1996). *Off track: When poor readers become "learning disabled."* Boulder, CO: Westview.
- Spires, H. A., & Donley, J. (1998). Prior knowledge activation: Inducing engagement with informational texts. *Journal of Educational Psychology, 90*(2), 249-260.
- Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review, 30*(3), 407-419.
- Stahl, S. A. (1998). Four questions about vocabulary knowledge and reading and some answers. In C. R. Hynd & S. A. Stahl (Eds.), *Learning from Text Across Conceptual Domains* (pp. 73-94). Mahwah, NJ: Erlbaum.
- Stahl, S. A., Hare, V. C., Sinatra, R., & Gregory, J. F. (1991). Defining the role of prior knowledge and vocabulary in reading comprehension: The retiring of number 41. *Journal of Reading Behavior, 23*(4), 487-508.
- Stahl, S. A., Jacobson, M. G., Davis, C. E., & Davis, R. L. (1989). Prior knowledge and difficult vocabulary in the comprehension of unfamiliar text. *Reading Research Quarterly, 31*(4) 430-456.
- Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.
- Stanovich, K. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core, variable-difference model. *Journal of Learning Disabilities, 32*, 590-612.

- Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, 20(1), 51-68.
- Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 24, 402-433.
- Stanovich, K. E., West, R. F. & Harrison, M. R. (1995). Knowledge growth and maintenance across the life-span: The role of print exposure. *Developmental Psychology*, 31(5), 811-826.
- Stevens, K. (1980). The effect of background knowledge on the reading comprehension of ninth graders. *Journal of Reading Behavior*, 12(2), 151-154.
- Stothard, S. E. (1994). The nature and treatment of reading comprehension difficulties in children. In C. Hulme & M. Snowling (Eds.), *Reading development and dyslexia* (pp. 200-238). London: Whurr Publishers.
- Symons, S., & Pressley, M. (1993). Prior knowledge affects text search success and extraction of information. *Reading Research Quarterly*, 28(3), 251-261.
- Tan, A., & Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. *Journal of Educational Psychology*, 89(2), 276-288.
- Taylor, B. M., & Beach, R. W. (1984). The effects of text structure instruction on middle-grade students' comprehension and production of expository text. *Reading Research Quarterly*, 19(2), 134-146.
- Tenenbaum, A. B. (1977). Task-dependent effects of organization and context upon comprehension of prose. *Journal of Educational Psychology*, 69(5), 528-536.

- Thorndike, R. L. (1973). *Reading comprehension education in fifteen countries*. New York: Wiley.
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S. & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second- to fifth-grade children. *Scientific Studies of Reading, 1*(2), 161-185.
- Trabasso, T., & Suh, S. (1993). Understanding text: Achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Processes, 16*, 3-34.
- Underwood, T., & Pearson, P. D. (2004). Teaching struggling adolescent readers to comprehend what they read. In T. L. Jetton & J. A. Dole (Eds.), *Adolescent literacy research and practice* (pp. 135-161). NY: Guilford.
- Urduan, T. (2004). Predictors of academic self-handicapping and achievement: Examining achievement goals, classroom goal structures, and culture. *Journal of Educational Psychology, 96*(2), 251-264.
- van den Broek, P., Lynch, J. S., Naslund, J., Ievers-Landis, C. E., & Verduin, K. (2003). The development of comprehension of main ideas in narratives: Evidence from the selection of titles. *Journal of Educational Psychology, 95*(4), 707-718.
- van den Broek, P., Risdien, K., Fletcher, C. R., & Thurlow, R. (1995). A “landscape” view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In B. K. Britton & A. Graesser (Eds.), *Models of understanding text* (pp. 165-187). Mahwah, NJ: Erlbaum.

- van den Broek, P., Tzeng, Y., Ridsen, K., Trabasso, T., & Basche, P. (2001). Inferential questioning: Effects on comprehension of narrative texts as a function of grade and timing. *Journal of Educational Psychology, 93*(3), 521-529.
- van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2004) Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology, 96*(1), 19-30.
- van Kraayenoord, C. E., & Schneider, W. E. (1999). Reading achievement, metacognition, reading self-concept and interest: A study of German students in grades 3 and 4. *European Journal of Psychology of Education, 14*(3), 305-324.
- VanSledright, B. A. (1995). "I don't remember—the ideas are all jumbled in my head": 8th graders' reconstructions of colonial American history. *Journal of Curriculum and Supervision, 10*(4), 317-345.
- Vidal-Abarca, E., Martínez, G., & Gilabert, R. (2000). Two procedures to improve instructional text: Effects on memory and learning. *Journal of Educational Psychology, 92*(1), 1-10.
- Viola, H. J., Wheatley, H., & Hart, D. (1998). *Why we remember*. Menlo Park, CA: Addison-Wesley.
- Wade, S. E., Trathen, W., & Schraw, G. (1990). An analysis of spontaneous study strategies. *Reading Research Quarterly, 25*(2), 147-166.
- Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction, 22*(3), 333-362.
- Wagner, M., & Rohwer, W. D. (1981). Age differences in the elaboration of inferences from text. *Journal of Educational Psychology, 73*(5), 728-735.

- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S., Barker, T. A., Burgess, S., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33*(3), 468-479.
- Walczyk, J. J. (1990). Relation among error detection, sentence verification, and low-level reading skills of fourth graders. *Journal of Educational Psychology, 82*(3), 491-497.
- Walczyk, J. J. (1994). The development of verbal efficiency, metacognitive strategies, and their interplay. *Educational Psychology Review, 6*(2), 173-189.
- Walczyk, J. J. (2000). The interplay between automatic and control processes in reading. *Reading Research Quarterly, 35*(4), 554-566.
- Walczyk, J. J., Marsiglia, C. S., Bryan, K. S., & Naquin, P. J. (2001). Overcoming inefficient reading skills. *Journal of Educational Psychology, 93*(4), 750-757.
- Walczyk, J. J., & Taylor, R. W. (1996). How do the efficiencies of reading subcomponents relate to looking back in text? *Journal of Educational Psychology, 88*(3), 537-545.
- Weinstein, C. E., Husman, J., & Dierking, D. R. (2000). Self-regulation interventions with a focus on learning strategies. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 727-747). San Diego: Academic Press.
- White, T. G., Graves, M. F., & Slater, W. H. (1990). Growth of reading vocabulary in diverse elementary schools: Decoding and word meaning. *Journal of Educational Psychology, 82*(2), 281-290.

- Whitney, P., & Budd, D. (1996). Think-aloud protocols and the study of comprehension. *Discourse Processes, 21*(3), 341-351.
- Whitney, P., Budd, D., Bramucci, R. S., & Crane, R. S. (1995). On babies, bath water, and schemata: A reconsideration of top-down processes in comprehension. Sentences. *Discourse Processes, 23*(3), 337-356.
- Wigfield, A. (1997). Reading motivation: A domain-specific approach to motivation. *Educational Psychologist, 32*(2), 59-68.
- Wiley, J. & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*(2), 301-311.
- Williams, J. P. (1993). Comprehension of students with and without learning disabilities: Identification of narrative themes and idiosyncratic text representations. *Journal of Educational Psychology, 85*(4), 631-641.
- Wilson, C. R., & Hammill, C. (1982). Inferencing and comprehension in ninth graders reading geography textbooks. *Journal of Reading, 24*, 424-428.
- Winne, P. (2001). Self-regulated learning viewed from models of information processing. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153-189). Mahwah, NJ: Erlbaum.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304) Mahwah, NJ: Erlbaum.

- Winne, P.H., Jamieson-Noel, D., & Muis, K. (2002). Methodological issues and advances in researching tactics, strategies, and self-regulated learning. In P. Pintrich & M. Maehr (Eds.), *New directions in measures and methods* (pp. 121-156). Amsterdam, The Netherlands: JAI.
- Wirt, J., Choy, S., Gerald, D., Provasnik, S., Rooney, P., Watanabe, S., & Tobin, R. (2002). *The condition of education 2002*. Washington, DC: U.S. Department of Education.
- Wixson, K. (1986). Vocabulary instruction and children's comprehension of basal stories. *Reading Research Quarterly*, 21(3), 317-329.
- Woodcock, R. W. (1997a). *Woodcock diagnostic reading battery*. Itasca, IL: Riverside.
- Woodcock, R. W. (1997b). *Examiner's manual for the Woodcock diagnostic reading battery*. Itasca, IL: Riverside.
- Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5(3), 161-215.
- Yopp, R. H., & Dreher, M. J. (1994). Effects of active comprehension instruction on attitudes and motivation in reading. *Reading Horizons*, 34, 288-302.
- Zwaan, R. A., & Brown, C. M. (1996). The influence of language proficiency and comprehension skill on situation-model construction. *Discourse Processes*, 21(3), 289-327.