UNDERSTANDING READING COMPREHENSION PERFORMANCE
IN HIGH SCHOOL STUDENTS

by

Bozena Kwiatkowska-White

A dissertation submitted to the Faculty of Education
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
August, 2012
copyright© Bozena Kwiatkowska-White 2012
Reading furnishes the mind only with materials of knowledge; it is thinking that makes what we read ours.

from *Of The Conduct Of The Understanding* by John Locke (1632-1704)
ABSTRACT

The ability to extract meaning from text is an important skill. Yet many students struggle with effectively comprehending what they read. In comparison with research carried out with younger students, there is a lack of research in the reading comprehension of adolescents (Grades 4 – 12). The goal of this dissertation was to increase our understanding of the factors that underlie the poor reading comprehension abilities of this older group of students. This dissertation includes two studies drawn from a sample of 137 age 15 year old high school students. Study One utilized archival data from government mandated tests of reading achievement of 78 students administered in Grades 3, 6, and 10, and results from a commercially available test of reading comprehension administered in Grade 10. This longitudinal study examined the prevalence of the stability, cumulative growth, and compensatory models in reading comprehension development. Probabilities of later-grade reading achievement categorization conditioned on earlier-grade reading achievement were computed, the prevalence of developmental paths was estimated, and tests of regression to the mean were conducted. Overall findings suggest considerable stability across time.

Study Two examined the specificity of the comprehension weaknesses of 15 year old readers whose comprehension skills are below those expected from their skill in word reading and nonverbal ability (unexpected poor comprehenders). Regression analyses identified unexpected poor comprehenders, and two contrast groups (expected average and unexpected good comprehenders). Characteristics of unexpected poor comprehenders are examined after controlling for word-reading accuracy, phonological decoding, reading rate, nonverbal ability, and vocabulary. Findings indicate a critical disadvantage of unexpected poor comprehenders lies in their weakness in vocabulary and that comprehension difficulties related to the identification of details and main ideas in summary writing remain when
vocabulary is controlled. Implications for interpreting previous and informing future research are discussed.

Results of both studies are discussed with respect to the nature of the reading comprehension construct, identification and remediation of reading comprehension difficulties, and the assessment of reading comprehension.
ACKNOWLEDGEMENTS

This thesis would not have been possible without the support of many people.

First, and foremost, I would like to express my gratitude to John Kirby, my supervisor, who allowed me to ‘jump from high bridges’. His support and guidance made it possible for me to not only complete this thesis, but also to develop a theoretically guided understanding for approaching the complexity of reading comprehension. His expertise in both the reading and comprehension components of reading comprehension have allowed me to bridge two fields that traditionally have been examined separately. He opened the door for me to work with teachers in improving literacy outcomes for adolescent students. He showed me the importance of acquiring both depth and breadth in my knowledge base.

To my committee members. Elizabeth Lee, who through her friendship, was always available to provide me with the encouragement and feedback necessary for me to grapple with this complex construct. To Nancy Hutchinson who willingly jumped in at a very late date as a committee member.

To Kate Cain, whose research and professional support provided me with a point of departure for approaching the study of unexpected poor comprehenders.

To my family members. Erin, as with many other things we have shared, we travelled this journey together and completed each of our theses at the same time. I look forward to continuing the sharing of knowledge and support as you embark on your new position at Baycrest. To my mother, who never doubted the value of my goals and achievements. She through her work at the University of Toronto provided me with much insight into the Ph.D. process. She repeatedly acknowledged the positive influence that John Kirby has had on my development. To Stephanie, who I know has always been proud of my accomplishments. To my sister, Grazyna, who has always been there for me, and whose keen intellect has always
been admired. To my niece, Rebecca, who as she was completing her thesis, taught me about the literal and metaphorical value of wearing a ‘hoodie’, or ‘cap’ on one’s head when writing.

To my valued friend, Mario, who provided all kinds of support during this journey.

To the administration, staff, and students at St. Theresa’s High School who provided me with the opportunity to collect my data.
TABLE OF CONTENTS

ABSTRACT........................................................................................................................................... ii

ACKNOWLEDGEMENTS ......................................................................................................................... iv

LIST OF TABLES...................................................................................................................................... x

LIST OF FIGURES.................................................................................................................................... xii

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW ................................................................. 1

Introduction ........................................................................................................................................... 1

What is Reading Comprehension? ........................................................................................................ 3

The Simple View of Reading ................................................................................................................ 3

Summary and Limitations .................................................................................................................... 7

The Role of Word Knowledge in Reading Comprehension ................................................................. 7

Importance of Word Reading Skill for Reading Comprehension ....................................................... 8

Processes on which Word Reading is Dependent ............................................................................. 9

The Lexical Quality Hypothesis ........................................................................................................... 10

Summary ............................................................................................................................................... 16

Construction-Integration Model of Comprehension .......................................................................... 16

Summary and Limitations .................................................................................................................... 24

Summary of Models and Theories ...................................................................................................... 24

Factors Contributing to Poor Comprehension ................................................................................. 26

Details of Dissertation Studies ........................................................................................................... 28

Study One: Development of Reading Comprehension .................................................................... 28

Study Two: Unexpected Poor Comprehenders ................................................................................. 29

References ............................................................................................................................................. 34

Chapter 2: A Longitudinal Study of Reading Comprehension Achievement from Grades 3 to 10:
Investigating Stability, Cumulative, and Compensatory Models .................................................... 45

Abstract ............................................................................................................................................... 45

Introduction .......................................................................................................................................... 45

Patterns of Reading Comprehension Growth .................................................................................... 46

The Stability Model ............................................................................................................................ 46

The Cumulative Model ....................................................................................................................... 49
LIST OF TABLES

Table 2.1 Descriptive Statistics (N=78) ........................................................................................................65
Table 2.2 Correlations between Reading Comprehension Measures .................................................65
Table 2.3 Conditional Probabilities of Reading Achievement Categorization in Grade 6
Conditioned on Grade 3 Achievement .................................................................................................67
Table 2.4 Conditional Probabilities of Grade 10 Reading Achievement Categorization
Conditioned on Grade 6 Achievement .................................................................................................68
Table 2.5 Conditional Probabilities of Grade 10 Reading Achievement Categorization
Conditioned on Grade 3 Achievement ...............................................................................................69
Table 2.6 Prevalence of Development Paths in Grade 6 and 10 Based on Grade 3 Reading
Achievement .........................................................................................................................................72
Table 2.7 Prevalence of Developmental Paths from Grade 3 and 6 Based on Grade 10
Reading Achievement .........................................................................................................................74
Table 2.8 Actual (A), Predicted (P), and Difference (D) Scores for each Time Period ............78
Table 3.1 Means and Standard Deviations of all Variables (N = 137) ..............................................116
Table 3.2 Correlations Among all Variables (N = 137) .................................................................117
Table 3.3 Means and Standard Deviations of Unexpected Poor, Expected Average, and
Unexpected Good Comprehenders on Predictor and Outcome Variables in Four Models ... 121
Table 3.4 Summary of Analyses of Variance Comparing Unexpected Poor, Expected
Average, and Unexpected Good Comprehenders on Control Variables for Four Models .... 123
Table 3.5 Consistency of Group Membership across Four Models (rows show percent of
students classified under each model of those classified in column model) ..................125
Table 3.6 Means and Standard Deviations for Individual Differences, Word Related, and
Reading Comprehension Variables of Unexpected Poor, Expected Average, and Unexpected
Good Comprehenders (Model 4) ........................................................................................................127
Table 3.7 ANOVA and ANCOVA Results for Comprehension-related Variables (Model 4)..........................131

Table 4.1 Achievement of Unexpected Poor, Expected Average, and Unexpected Good Comprehenders in Grades 3, 6, and 10.................................................................151

Table F.1 Propositional template for Summary Writing Task.............................................190
LIST OF FIGURES

Figure 2.1  Comparison of Time-forward (solid lines) with Time-backwards Analysis (broken lines) for each Achievement Group for Grade 3 and 6, Grade 6 and 10, Grade 3 and 10 ................................................................. 76

Figure 2.2  Progress of Children based on Actual (solid lines) and Predicted Scores (dotted lines) of Achievement Groupings for each Time Period ................................................................. 78

Figure 3.1  Scatter Plot for Selection of Unexpected Poor, Expected Average, and Unexpected Good Comprehenders in Model 4 (word reading accuracy, reading rate, nonverbal ability, and vocabulary) ................................................................. 118
CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

Introduction

The ability to read and to extract meaning from text is a fundamental skill necessary for most forms of personal learning, intellectual growth, and educational attainment (United Nations Educational, Scientific and Cultural Organization, 2010). On an international level, the literacy rate in Canada is high: results from international assessments of reading proficiency indicate that Canadian youths (age 15) rank in third position of 41 countries (Bussière, Cartwright, & Knighton, 2004). Although the reading proficiency of Canadian youths places Canada as one of the top ranked countries in the world, the results mask a significant problem that demands attention: almost 30% of 15-year-old Canadian students did not meet what are considered to be adequate levels of proficiency on international tests of reading (OECD, 2010). That a substantial number of youth struggle to understand what they have read carries negative social and economic costs: poor youth literacy is related to high school dropout rates, long-term unemployment, and higher crime rates (The Canadian Language and Literacy Research Network, 2009).

The importance of monitoring the reading proficiency of youth is evident in the increased use of large-scale government-mandated assessment. In Ontario, the Education Quality and Accountability Office (EQAO) is responsible for administering tests assessing the reading proficiency of children when they are in Grades 3, 6, and 10. Results from these assessments also indicate that a substantial portion of students in Ontario have difficulty achieving an adequate level of reading proficiency (Grade 3, 39%; Grade 6, 31%, Grade 10, 15%; EQAO, 2010). Findings from these and international assessments of reading proficiency are limited in that they do not provide information regarding (a) the changes in reading achievement that individuals may experience over time; and, (b) the possible sources of difficulties for those students who struggle with effective reading comprehension. Studies
that address these issues are necessary to inform our theoretical understanding of the reading comprehension construct, what effective reading comprehension depends upon, and how it develops. They are also necessary so that appropriate remediation can be targeted at the earliest time possible in the development of this important skill.

Although great progress has been made in understanding learning to read in young children (ages 5-8) (e.g., National Reading Panel, 2000), far less is known about reading in older children, or about comprehension (reading to learn). What is known is far from sufficient to enable us to instruct students optimally, much less intervene with those who struggle to comprehend what they read (Shuy, McCardle, & Albro, 2006). The lack of basic research in the reading comprehension of adolescents (Grades 4 – 12) has been clearly identified in a number of documents (Berman & Biancarosa, 2005; Carnegie Council on Advancing Adolescent Literacy, 2010; National Institute of Child Health and Human Development, 2000; RAND Reading for Understanding Report, 2002). The goal of this dissertation is to increase our understanding of the factors that underlie the poor reading comprehension abilities of this older group of students. The ultimate aim of the dissertation is to improve our understanding of reading achievement, and to provide better information for those aiming to improve reading comprehension in adolescent students.

Below, I define what reading comprehension is and review three influential models or theories that have been used to study the reading comprehension construct. These include the simple view of reading, the lexical quality hypothesis, and the construction-integration model. I will then show how my research fits into these models and theories.

My dissertation is comprised of two inter-related studies carried out with 137 Grade 10 (15 year old) students, that utilized archival government reading achievement data collected from when these students were in Grades 3, 6, and 10. The primary aim of Study One was to examine the development of reading comprehension achievement over time in
order to understand whether individual differences in reading ability persist over time. The primary aim of Study Two was to examine both lower level word reading skills and higher level cognitive processes that have been related to comprehension in order to understand more about their relationship in reading failure. Following Study Two, I provide a short chapter relating the results of the two studies.

What is Reading Comprehension?

Reading comprehension is “the process of simultaneously extracting and constructing meaning through interaction and involvement with written language” (RAND Reading Study Group, 2002, p.11). This process of interaction and involvement with the text is a function of both reader and text variables that take place within a larger social context (Goldman, Saul, & Coté, 1995; McNamara & Magliano, 2009; RAND Reading Study Group, 2002). When successful, the product of reading comprehension is a coherent mental representation of a text’s meaning that is integrated with the reader’s prior knowledge. This product is often referred to as a mental model (Johnson-Laird, 1983) or a situation model (Kintsch, 1998; Kintsch, & van Dijk, 1978) and is considered to be the basis for learning from text. The nature of the model, that is the ideas and the links connecting those ideas, defines what has been learned.

Reading comprehension is a complex skill: it requires the successful development and orchestration of a variety of lower- and higher-level processes and skills (Balota, Flores d’Arcais, & Rayner, 1990). As a consequence, there are a number of sources for potential comprehension failure and these sources can vary depending on the skill level and age of the reader (Keenan, Betjemann, & Olson, 2008; RAND, 2002). Theories and models of reading comprehension are necessary to make sense of this complexity.

The Simple View of Reading

The simple view of reading (Gough & Tunmer, 1986; Hoover & Gough, 1990) proposes that reading comprehension (RC) comprises the product (RC = D x LC) of two
component skills: decoding (D) and listening comprehension (LC). According to this view, the D component translates print into language, and the comprehension component of LC makes sense of this linguistic information. From this perspective, the development of reading comprehension skill requires that the reader develop skill in both these components.

Empirical findings have supported the simple view, and its ability to account for a substantial amount of the variance in reading comprehension performance (e.g., Cutting & Scarborough, 2006; Hoover & Gough, 1990). For example, it has been found to account for approximately 40% to 80% of the variance in reading comprehension with readers ranging from 8 to 16 years old (Catts, Adlof, Hogan, & Ellis Weismer, 2005; Dreyer & Katz, 1992; Johnston & Kirby, 2006; Joshi & Aaron, 2000; Savage, 2006).

One of the major benefits of the simple view is that it can be used as a tool for classifying children’s reading abilities and impairments (e.g., Catts, Hogen, & Fey, 2003). Whereas good text comprehenders have been generally found to have strengths in both D and LC, poor text comprehenders have been shown to have deficits in: (a) the D component (e.g., Catts et al., 2006; Shankweiller, Lundquist, Katz, Stuebing, Fletcher, Brady, & Shaywitz, 1999), (b) the LC component (e.g., Cain & Oakhill, 2007; Catts et al., 2003), (c) or both (e.g., Kamhi & Catts, 2002). Such a classification system is valuable for ensuring that remediation is targeted at the specific source(s) of an individual’s reading comprehension difficulties (Nation, Snowling, & Clarke, 2007).

One issue that has been raised in the literature pertaining to the simple view of reading relates to the contribution that each component makes to reading comprehension ability and how this relationship might change as reading ability develops. Some researchers have found that the relationship between D and RC weakens in the later grades, once decoding skills are mastered, and that then the strength of the relationship between listening comprehension and reading comprehension increases (Hoover & Gough, 1990; Stanovich, 1991; Vellutino,
Scanlon, Small, & Tanzman, 1991). Other researchers, while concurring with the latter finding, have found evidence that the relationship between D and RC does not weaken as grade level increases (Carver, 1998; Dreyer & Katz, 1992). Sample-specific differences in component growth rates may account for the differences between studies (Johnson & Kirby, 2006), but another explanation may be derived from inconsistencies in how the D construct is defined and measured (Carver, 1998).

One ambiguity within the simple view model is whether decoding refers to successful word reading (timed or untimed), or instead refers to the ability to use phonetic analysis (as indexed by pseudoword reading) (see Johnston & Kirby, 2006; Savage, 2006). This distinction is important to our understanding of the relative contribution that D makes to RC, as a greater number of sub-processes (reviewed below) are implicated in the former, thereby making a larger overall contribution to RC than does the latter (Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Johnson & Kirby, 2006).

Another explanation for the observed differences in the contributions made by D and LC to RC comes from the findings of studies that have shown that the correlations of D and LC with RC depend upon which measure of RC is used. Spear-Swerling (2004) observed that word reading accuracy accounted for more of the variance in a cloze-test measure of RC (51.3%) than in one that used a short answer format (33.5%). Cutting and Scarborough (2006) found that the same set of predictors (i.e., word reading, oral language measures) predicted performance on different standardized assessments of reading comprehension differently. In their analyses, they included three measures of reading comprehension commonly used by researchers: the Weschler Individual Achievement Test (WIAT; Wechsler, 1992), the Gates-MacGinitie Reading Test-Revised (G-M; MacGinitie, MacGinitie, Maria, & Dreyer, 2000) and the Gray Oral Reading Test-Third Edition (GORT-3; Wiederhold & Bryant, 1992). The G-M and WIAT correlated more strongly with each other ($r = .79$) than they did with the
GORT-3 \( (r = .64 \text{ with } \text{G-M}; \ r = .70 \text{ with } \text{WIAT}) \). Cutting and Scarborough also reported that almost twice as much unique variance was accounted for by word reading measures (composite of word reading and decoding) on the WIAT (11.9%) than on G-M (6.1%) and GORT (7.5%). Differences in the unique contributions of the language composite (composed of sentence comprehension, syntax, and vocabulary) were also noted (WIAT and GORT 9%; G-M 15%). Cutting and Scarborough also found differences in the combined contributions of decoding and oral language measures to the different reading comprehension measures. Whereas a substantial portion of the variance on the G-M and WIAT (67% and 72% respectively) was accounted for by their D and LC measures, a much smaller portion was accounted for by them on the GORT (49%). The assumption that different tests used to assess reading comprehension are comparable appears questionable (Keenan, Betjemann, & Olson, 2006). This raises the question of what comprehension in either the LC or RC terms constitutes. This does not appear to be an issue that has been well addressed in the simple view of reading literature from neither a theoretical, nor an empirical perspective.

Findings that different reading comprehension measures tap different cognitive processes (Cutting & Scarborough, 2006; Keenan et al., 2006) reinforce the importance of using parallel tests when assessing LC and RC (Hoover & Gough, 1990). But, as Kirby and Savage (2008) point out, there is a crucial difference between the RC and LC situations: in the former the content stays present whereas in the latter it disappears. This difference raises the question as to whether other factors, such as metacognitive strategies\(^1\), may affect

\(^1\) Metacognition refers to individuals’ awareness of their cognitive processes and strategies (Flavell, 1979; Flavell, Green, & Flavell, 1995). Examples of metacognitive strategies include the use of background knowledge, drawing inferences, question generation, note taking and summarizing; strategies that research has found receptive to teaching and that when used effectively have been found to improve comprehension (see Jones, 2007 and National Reading Panel, 2000, for reviews relating to listening and reading comprehension, respectively).
comprehension differently depending upon whether comprehension is assessed within a listening, or a reading context. Kirby and Savage suggested that the RC situation leaves more scope for the application of metacognitive strategies than the LC situation:

If one has not found a main idea after reading a paragraph, one can go back and look again; this is much more difficult when one has only heard the text. Strategies such as locating information, finding main ideas, determining text structure, and using headings and typographical cues are either more difficult or impossible in LC than in RC. RC allows more scope for metacognition, because the metacognition does not interfere with reception of the content. It is possible that these deliberate, taught strategies either bring something to RC that is not present in LC, or even that they can help compensate for lower levels of LC. (pp. 79-80).

The authors conclude that further research is needed to examine how strategies affect a full range of reading comprehension measures within the simple view model.

**Summary and Limitations**

The simple view of reading provides a general framework for understanding reading comprehension in that it identifies two main components of reading comprehension: decoding and listening comprehension. The distinct patterns that children exhibit in these two components are useful for identifying the source of reading comprehension difficulties and appropriately targeting remediation efforts (Catts, Adlof, & Ellis Weismer, 2006; Nation et al., 2007). The simple view of reading, however, is limited in that it is neither a theory of word reading nor of comprehension. As Kirby and Savage (2008) pointed out in their review of the simple view of reading model, there is a need from both theoretical and applied perspectives to explicate more clearly each of these components.

**The Role of Word Knowledge in Reading Comprehension**

Theories of reading comprehension, such as the Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2001, 2002), propose that some, but not all, reading comprehension difficulties arise from weaknesses at the word level. Before turning to a review of this theory, the empirical findings that support it, and its limitations, I review the (a) empirical evidence
for the importance of word reading skill for reading comprehension, and (b) processes upon which an extensive body of research has found word reading to be dependent.

**Importance of Word Reading Skill for Reading Comprehension**

In the reading literature, the strongest and most replicated empirical support for the importance of skill in word reading for effective reading comprehension comes from two findings. First, children (and adults) with reading disabilities have deficits in word decoding (accurately attaching a name to a written word or pseudo-word) (see Vellutino et al., 2004 for a review). These deficits have been observed, often in the absence of any other reading-specific or general deficits (i.e., orthographic skill, verbal IQ; Felton & Wood, 1992; Osmon, Braun, & Plambeck, 2005). Second, there are consistent findings of positive and significant correlations between word recognition and reading comprehension (see Slocum, Street, & Gilberts, 1995, for a review). Skills in decoding and reading comprehension continue to be related even when decoding skill has become more automatized (Miller-Shaul, 2005; Vellutino & Scanlon, 1988). For example, correlations between the Woodcock-Johnson Word Identification and Passage Comprehension subtests are .82, .67, .59, .57, and .45 for grades 1, 3, 5, 8, and 11 respectively. These associations have been found across writing systems, particularly those with deeper orthographies (i.e., English versus Finnish) where better decoding skills are necessary to navigate less consistent symbol to sound associations (see Aaron & Joshi, 1989 and Perfetti & Bolger, 2004, for reviews). Importantly, correlations between measures of decoding and reading comprehension have been found to be stronger than between decoding and listening comprehension (Macaruso & Shankweiller, 2010; Tilstra, McMaster, Van den Broek, Kendeou, P., Rapp, 2009), and decoding measures have been found to account for unique variance in reading comprehension above that accounted for by listening comprehension (Bell & Perfetti, 1993; Scarborough & Cutting, 2006).
Although these associations provide support for the importance of word reading skill for effective reading comprehension, they are limited in that they do not provide us with a clear understanding of the processes on which word reading skill is dependent, or how word reading affects comprehension. Explicating these processes is a crucial component for not only advancing our theoretical knowledge base, but also for appropriately informing the teaching of effective word reading skill.

**Processes on which Word Reading is Dependent**

Word reading can be decomposed into its component processes (see Kirby, 2007). At the first level below word reading, there are phonological decoding and orthographic processing skills and knowledge. These two processes correspond to the two paths of the *dual route model* of reading (Castles, 2006). As children acquire facility in applying their knowledge of the relationship between graphemes and phonemes (phonological route), the role of orthographic processing becomes more important for effective word recognition. In their *grain size hypothesis*, Ziegler and Goswami (2005) argued that through reading experience, larger units (rimes, or even entire words) rather than individual letters, are processed via the orthographic route. Further, in a script like English where the correspondence between graphemes and phonemes is not always predictable, readers rely less on the phonological route and to a greater extent on the orthographic route (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Katz & Frost, 1992). Underlying the phonological and orthographic paths are other lexical and sublexical processes. For example, phonological decoding has been shown to depend upon phonological awareness (Vellutino et al., 2004). Skill in orthographic processing, on the other hand, depends upon the amount of reading experience (exposure to print) (Gustafson, 2001). Some researchers also suggest that effective word reading depends upon Rapid Automatized Naming (RAN) (Kirby, Parrila, & Pfeiffer, 2003).
A third cognitive process that influences word recognition is semantic knowledge. It is represented when words are read in context (e.g., Rayner & Well, 1996), when the meaning of the word is known (Adams, 1990), and in morphology (e.g., Carlisle, 2003). For example, knowing that the context is about Christmas may speed up recognition of the word *wreath*. Being aware of morphology (meaning-bearing units in words) may speed recognition of a morphologically complex word such as *unattached* (Carlisle & Stone, 2005). In sum, there is considerable evidence that word recognition is influenced by phonology, orthography, and semantics.

**The Lexical Quality Hypothesis**

The lexical quality hypothesis (Perfetti & Hart, 2001, 2002) attempts to outline the qualities of word representations that support effective reading comprehension. Lexical representations are the word entries in the lexicon, or mental dictionary, acquired through spoken or written language. According to the lexical quality hypothesis, it is the strength and stability of word entries in the lexicon that define their lexical quality (Perfetti, 2007). The lexical quality hypothesis was built on modifications outlined in Perfetti’s verbal efficiency theory (Perfetti, 1985) and the ideas outlined by him in the representation problem in reading acquisition (Perfetti, 1992).

In verbal efficiency theory, the rapid retrieval of a word’s phonology and meaning were considered to be a limiting factor in comprehension. Although two components (phonology and semantics) were identified as necessary for lexical representations, it was the phonological processes that allowed a word, or nonword, to be decoded, regardless of whether or not meaning was retrieved (Perfetti, 1985; see also Frost, 1998, for a review of studies supporting a strong phonological theory in visual word recognition). As such, the theory assumed that the ability to decode nonwords was the hallmark of basic alphabetic reading skill (Perfetti, 2007). The link from word-level reading to comprehension was
through the assumption that comprehension included higher level processes (e.g., integration and inferencing), that, in turn, required cognitive resources (e.g., working memory). Efficient and automatic word recognition were considered to free resources, which could then be devoted to higher-level linguistic processing, such as parsing, deriving sentence meanings or forming inferences across sentences (Bell & Perfetti, 1994; Perfetti, 1985, 1992; Perfetti & Hogaboam, 1975).

In summary, verbal efficiency emphasizes decoding, phonological processes, retrieval, memory, automaticity. Whereas the processing mechanisms in skilled reading are more efficient, in less skilled reading the same mechanisms operate less efficiently. However, as Perfetti himself has admitted, this view does not address the source of these differences in efficiency (Perfetti, 2010). And, as demonstrated in Fredericksen’s (1981) studies with high school students, improvements in speed and accuracy of word reading do not necessarily translate into improvements in reading comprehension. Further, the general observation of reduced correlations between word recognition and reading comprehension in older and more skilled readers suggests that there is plenty of room for disassociation between the two (Catts, et al., 2005; Landi, 2010; Leach, Scarborough, & Risco, 2003).

Perfetti (1992) revised his definition of lexical quality in the representation problem in reading acquisition to include two new concepts: specificity and redundancy. Lexical representations need to be specific; a single orthographic representation must be associated with the appropriate lexical entry. For example, the word *bank* has low specificity because one orthographic representation is associated with two meanings (*river edge* and *financial institution*). Lexical representations also need to be built and accessed with redundant (duplicate) information; sound information from speech and print must be associated with the same lexical entry. For example, the word *sow* has low redundancy because a single orthographic representation is associated with two sound representations and two meanings
(female pig and plant crop). The concepts of specificity and redundancy have consequences for reading ability and its acquisition. Experimental studies have shown slowed response times for homographs compared to control words with a rise in activation for both lexical items (Gottlob, Goldinger, Stone, & Van Orden, 1999; Hart & Perfetti, 2008). The competition between two lexical items results in a period of indecision evidenced by longer response times. Differences in response time patterns between poor readers and fluent readers can indicate processes contributing to the reading difficulties of the poor group (Gernsbacher & Faust, 1991; Perfetti & Hart, 2001, 2002). The two concepts of specificity and redundancy also emphasize that word representations are acquired incrementally rather than in discrete stages.

Built on modifications of verbal efficiency theory and the representation problem in reading acquisition, the lexical quality hypothesis (Perfetti & Hart, 2001, 2002) applies a triangular model utilized by connectionist models of reading (e.g., Plaut, 2005) and outlines three constituents, phonology, orthography, and semantics, that underlie word representations. The theory claims that variation in the quality of word representations has consequences for reading skill, including comprehension (Perfetti, 2007). Individual differences in reading comprehension can be understood by the degree to which the constituents vary in quality, and hence the unity and completeness of a lexical representation. Quality is the extent to which a mental representation of a word specifies its constituents of form (phonology and orthography) and meaning (semantics) in a way that is both precise and flexible (Perfetti, 2007). The precision is needed because gate and gait, although phonologically the same, differ in their orthographic and semantic constituents. Both precision and flexibility are needed to understand and correctly pronounce record in “You need a record of the purchase” and “They can’t record the song”. If word codes for all three constituents are of high quality, semantic activations will occur more quickly, and deactivations for associated word codes
that are not relevant in a given context will occur earlier than if they are of low quality (Perfetti, 2001). The more coherent (complete, or bound versus fragmented or incomplete) the three constituents are, the more efficient the word identification system is, making more resources available for combining identified words into a meaningful representation of the text’s message. In short, the emphasis in the lexical quality hypothesis is on word knowledge comprising both form and meaning.

**Empirical support.** Some empirical support for the lexical quality hypothesis has been provided. For example, Perfetti and Hart (2001, 2002) explored the factor structure of word knowledge and comprehension skill in college-aged skilled and less-skilled readers. When accuracy measures were used, skilled readers’ knowledge of spelling, phonology, and decoding could be represented reasonably well by a word-form factor and a second factor that was defined by meaning and comprehension. Word identification loaded on both these factors. In contrast, less skilled readers, in addition to a meaning factor, required two form factors - one was defined more by phonological tasks and the other by orthographic tasks, suggesting less coherence of word identities. When speed measures were used, the factor structures that emerged again suggested a more coherent lexical knowledge structure for skilled readers than for less-skilled comprehenders. For skilled readers, vocabulary knowledge was highly related to knowledge of word form (spelling and pronunciation), loading on a single factor. For less-skilled comprehenders, this relationship was much weaker - vocabulary knowledge loaded on a factor that was separated from the form knowledge factor. Based on correlational evidence, their results point to qualitative differences in lexical representations for strong and weak readers, with a stronger coherence among different dimensions of word knowledge for more-skilled readers.

**Limitations.** An important issue that is not well addressed in the lexical quality hypothesis concerns an adequate account of the factors and processes that contribute to
individual differences in the lexical quality of word representations. Perfetti and Hart (2001, 2002) suggested that less-skilled comprehenders lack sufficient experience in reading that can help to establish new form–meaning relationships. Indeed, Perfetti has argued that it is multiple and effective experiences with words that lead to an abstracted representation of the form (phonology, orthography) and meaning of a word (Frishkoff, Collins-Thompson, Perfetti, & Callan, 2008; Landi & Perfetti, 2006). Indirect support for this comes from studies that have reported more skilled comprehenders to have greater reading experience than less skilled comprehenders (e.g., Perfetti & Hart, 2001; 2002; Landi, 2010; Stanovich, 1986). More direct support comes, for example, from a study by Bolger, Balass, Landen, and Perfetti, (2008). In this study, the effects of context variation in the learning of new words were tested by presenting adult learners with context sentences that either varied or repeated across learning trials. They found that experiencing new words in a variety of contexts led to better learning of their abstract meanings and orthographic forms compared with a single repeated context.

Although individual differences in reading experience may be one way of accounting for the development of high level representations of words, this account neglects the influence that other processes and factors may exert. A study by Perfetti, Wlotko, and Hart (2005) serves to illustrate this point. In this study, undergraduate students were given 45 minutes to learn the meanings of 60 very rare words (e.g., gloaming). The learning phase consisted of the presentation of the rare word on one side of a card and a brief definition on the opposite side. The effects of this learning were then tested using a meaning judgement task. Skilled comprehenders were reliably more accurate than less-skilled comprehenders. The authors provided three possible interpretations for the better word learning by the skilled comprehender group: (a) skilled comprehenders, who also have larger vocabularies, may be better able to use their word knowledge to add new words to their vocabulary; (b) skilled
comprehenders are better at learning new associations, or (c) skilled comprehenders are better at retaining specific episodic information. Importantly, what is missing from the lexical quality hypothesis, or any of the word representation theories, is an adequate account of how more skilled readers combine words in a text that will ultimately lead to a meaningful representation of the words that they have read, how new knowledge structures in long-term memory are built, or how the context in which a word appears exerts control over the information evoked from memory (see McKoon & Ratcliff, 1998 for a discussion). For example, with regard to the third issue, which members of a category are evoked by a category label (e.g., animal) depends on the sentence context in which the label is given (e.g., riding an animal versus milking an animal: Roth & Shoben, 1983).

Another issue that is not clearly articulated in the lexical quality hypothesis concerns the role of morphology in the development of word knowledge, and in reading comprehension performance. As noted earlier, morphology concerns the meaning-bearing units of word. More specifically, it is the conventional system by which the smallest units of meaning, called morphemes (bases, prefixes, and suffixes), combine to form complex words. A growing body of empirical research demonstrates the importance of morphological knowledge in improving reading outcomes (see Bowers, Kirby, & Deacon, 2010, for a review of studies). How, then, might morphological knowledge be reconciled within the lexical quality hypothesis? One way that has been suggested is to conceptualize morphological knowledge as a mechanism operating at the sub-lexical level that strengthens learners’ lexical representations (Carlisle & Katz, 2006; Carlisle & Stone, 2005; Bowers et al., 2010). Bowers and colleagues take this further and suggest that it is the feature of constituent binding, outlined in the lexical quality hypothesis, that pulls together individual features (phonology, orthography and semantics which includes morphology) of lexical representation that enhances lexical quality. Empirical support for this claim was provided by the findings of
their review of studies investigating the effects of morphology instruction. Specifically, they reported that increasing sublexical morphological knowledge through instruction resulted in improved scores on lexical measures. These findings suggest that individual differences in morphological knowledge, in addition to orthographic knowledge, phonological processing, and vocabulary, should be considered as a factor contributing to word knowledge. In order to examine the role that word knowledge plays in effective reading comprehension, a measure assessing each of these components is included in Study Two of my dissertation.

**Summary**

Theories of reading comprehension that focus on the role that word knowledge plays in effective reading comprehension are a useful starting point for improving our understanding of reading comprehension performance. However, these theories do not provide us with the whole story. To better assist us in understanding the nature and processes underlying effective reading comprehension, a more detailed theory of comprehension is necessary.

**Construction-Integration Model of Comprehension**

Kintsch’s (1988, 1998) construction-integration model provides a general processing framework for discourse comprehension. It offers an account of the processes involved in comprehension, and identifies the nature of the products that result from comprehension. His model, however, does not address word identification or any of the lower level factors of word knowledge (i.e., phonology, orthography, vocabulary) that are outlined in the lexical quality hypothesis. Rather, the processes of comprehension outlined in Kintsch’s model assume skilled word reading and begin after decoding and the semantic processes of attaching relevant word meanings have been successfully completed.

Development of Kintsch’s model was motivated by the limitations posed by previous schema-driven accounts that were used to describe the structure and organization of linguistic and discourse knowledge (Kintsch, 1988). Schema-based accounts of comprehension were
helpful in that they stressed the constructive nature of the reading process with an emphasis on the reader’s background knowledge, as well as the critical role of the interaction between the text and the reader’s background knowledge (e.g., Minsky, 1975; Schank, 1980; Schank & Abelson, 1977). However, schema-based accounts, with their underlying notions of top-down effects and expectation-driven processing, were considered to be too rigid and failed to provide an adequate account of how comprehension occurs in new contexts (Kintsch, 1988; Kintsch et al., 1993).

In Kintsch’s model, prior knowledge is considered to play a crucial role in how a text is interpreted, but in skilled comprehension the top-down processes are not considered to dominate; rather the text constrains and guides comprehension. Kintsch’s model, therefore, replaces the notion of pre-stored knowledge structures reflected in schema based theories with a minimally organized knowledge system that is generated in the context of the task for which it is needed. From this perspective, Kintsch’s construction-integration model describes the interplay between top-down (memory and knowledge) and bottom-processes (sensory input) in comprehension: how top-down processes guide comprehension and how bottom-up processes constrain it (Kintsch, 2005). In Kintsch’s model, the bottom-up processes do not include lower level word identification processes; rather, bottom-up processes begin after the propositional level of text meaning has been constructed (see McNamara & Magliano, 2009, for a discussion).

An account of the processes that are involved in comprehension is conveyed in the name of the model, the construction-integration model of comprehension. Construction refers to the activation of related knowledge (following principles of priming), including knowledge that is both relevant and irrelevant to the immediate or intended context. Potential sources of activation include the current input (i.e., word, proposition, and sentence), previous input, related knowledge, and reinstatements from prior text. These activation sources are often
represented as layers in a network of nodes and links. Whereas nodes are meant to represent words, propositions (idea units), or concepts, links are used to represent relationships between nodes. This knowledge activation process is often referred to as *dumb activation* because it assumes that top-down processing does not constrain the initial activation of knowledge. It has also been described as *retrieval based*, emphasizing the role of automatic memory retrieval process in comprehension (Kintsch, 1998, p.97). Activation sources from the construction process are then, in a second phase, integrated using a *constraint satisfaction* process: that is, the final mental representation is constrained by the degree to which concepts and the relations between concepts provided by the input and long-term memory have been activated. This process results in greater activation for concepts that are linked to other related concepts and a loss in activation for peripheral concepts that have fewer connections to other concepts in the mental representation. Only those concepts and ideas that are connected to many other concepts are left, whereas less connected concepts lose activation.

A key aspect of Kintsch’s model is its ability to account for the large demands that text comprehension places on working memory capacity. Kintsch’s model extends the traditional models of working memory that involve temporary storage (e.g., Baddeley, 2003). He adopted what is termed the *long-term working memory* perspective: recently activated or highly familiar information is quickly available from long-term memory by means of retrieval cues in short-term working memory. Because prior knowledge is considered to be continuously activated during comprehension, working memory capacity can be extended when the reader possesses a high degree of domain knowledge (Ericsson & Kintsch, 1995).

In addition to providing an account of the processes involved in comprehension, Kintsch’s model outlines three distinct but interacting levels of memory representations that are formed during comprehension (whether in listening or reading). When reading a text, the first level involves comprehension of the text’s surface structure. It is accomplished through
decoding processes and syntactic analysis. Although there are some cases when our goal in reading is to retain a verbatim representation of a text (e.g., memorizing a poem), when we read for understanding, the surface representation is typically not enduring. Rather, the surface representation is a mental representation consisting of a sequence of idea units or propositions. It is these propositions that serve as the basis for the reader to integrate the meanings of the individual clauses and sentences of a text. At this second level of comprehension, the idea units are interrelated into a network, called the microstructure of the text. The hierarchy of relations among the various sections of a text is also determined at this level, and is referred to as the macrostructure. Together the microstructure and macrostructure form what is termed the textbase. Textbase representations are considered to be relatively faithful to the presented passage. They consist of propositions derived from the input sentences of the text, plus inferences that are needed for referential coherence (Kintsch, 1998; p. 75).

The next level is considered to be the critical step in comprehension. It involves the construction of mental models (both visual and propositional in nature). In Kintsch’s model, this level is referred to as the situation model. It is a representation of the world that is reflected in the text and its relationship to the learner’s prior knowledge about the world. Similar to early schema theories, the situation model conveys the crucial notion that comprehension is more than deriving relationships between the words and ideas explicitly mentioned in a text. However, unlike schema theories, the situation model reflects the process whereby new knowledge is constructed by the reader. When readers create a situation model, they must generate inferences that go beyond the concepts that are explicitly mentioned in the text (text based inferences). These inferences must incorporate relevant background knowledge into the mental representation (knowledge based inferences). Comprehension at
this deeper level is considered to serve as the basis for learning – updating and modifying what an individual knows.

It is important to note that the situation model and the textbase are not entirely separate mental representations of the text content; rather they are viewed as different dimensions of the episodic memory of a text (Kintsch, 1988; van Dijk & Kintsch, 1983). Underlying the notion of hierarchical structure in these levels of representations is the assumption that processing proceeds from lower to higher levels of processing—that is the local meaning is computed prior to incorporating it into a more global text representation (Kintsch & van Dijk, 1978; Mitchell & Green, 1978). Therefore, a reader’s ability to form a coherent representation of the textbase, although necessary, is insufficient to ensure that a reader will engage in the inferential processes that support the construction of a coherent situation model (Kintsch, 1998; Kintsch & Kintsch, 2005).

A central issue in discourse comprehension research has been the extent to which comprehension is supported by bottom-up (data-driven) processes as compared to top-down (strategic) processes (e.g., Albrecht & Myers, 1995; Magliano & Radvansky, 2001; McNamara & Magliano, 2009; Myers & O’Brien, 1998, Tapiero, 2007). To this end, Kintsch made a distinction between two types of processes, passive versus active, which are used when building a mental representation of the text. When the reader (a) has the requisite prior knowledge (word and domain knowledge), (b) has an understanding of the goal state of comprehension, and (c) automatically and reliably generates the necessary retrieval structures that link the text information with relevant prior knowledge, comprehension will proceed smoothly and automatically at all levels. In this case, the reader relies on passive processes – without cognitive awareness, control, intention or effort – that demand few mental resources. It does not require the reader to build new knowledge structures in long-term memory. When the reader does not possess the pre-conditions (i.e., requisite prior knowledge, understanding
of the goal state, etc.), comprehension can break down. It is here that Kintsch highlights the relationship between comprehension and problem solving. Problem solving requires active processing. It is conscious and resource demanding, and when successful builds new knowledge structures. Active processing directed at lower levels of comprehension (propositional, text base), however, does not imply that a level of deeper comprehension (situation model) will be achieved (Kintsch & Kintsch, 2005).

One way of describing the relationship between Kintsch’s levels of memory representations and active processing is as depth of processing (Kirby & Woodhouse, 1994). The concept of depth of processing emphasizes how qualitative differences in a readers’ approach to learning (comprising motives and strategies) can lead to different comprehension outcomes (Biggs, 1987; Entwistle & Ramsden, 1983; Marton & Säljö, 1976). Students who have a deep approach to learning tend to be intrinsically motivated (they aim to understand) and employ meaningful learning strategies, such as integrating text material with their prior knowledge, resulting in a more active, deeper processing of the text. This can occur when the reader focuses on the main ideas, attempts to connect new information with prior knowledge, and derives a thematic understanding of the information in the text (Kirby & Woodhouse). Students who have a surface approach to learning, on the other hand, tend to be extrinsically motivated (they aim to do as little as possible) and employ superficial or rote learning strategies, such as breaking knowledge up into discrete fragments, and rote memorizing of those fragments, resulting in a more passive, superficial processing of the text (Kirby & Pedwell, 1991; Stein & Kirby, 1992).

Kirby & Woodhouse’s (1994) depth of processing is conceptually related to Kintsch’s construction-integration model in the following ways. First, depth of processing describes levels of reading comprehension outcomes (i.e., details, main ideas or themes) that are intended to reflect Kintsch’s micropropositions (for the surface level representation),
macropropositions (for the textbase) and situation model respectively. For example, when readers display comprehension at the thematic level they have formed a relationship among the textbase components in a way that is abstracted from the specific text context and have been able to comment on that relationship in some way (Williams et al., 2001, 2002). The thematic level of comprehension is conceptually related to Kintsch’s situation model in that both require the reader to form a representation of the world that is reflected in the text and its relationship to the reader’s prior knowledge about the world. Second, the distinction that depth of processing makes between two kinds of approaches readers have to learning (deep and surface) can be conceptually related to the distinction that Kintsch makes between active and passive processing. Both models emphasize how conscious and effortful processing is necessary for higher levels of comprehension as opposed to a reliance on superficial and effortless processing.

Although omitted in the construction-integration model, Kintsch has also emphasized the important role that different comprehension tasks play in promoting higher levels of comprehension (Kintsch, 1998, 2005). Tasks geared at surface levels of comprehension (e.g., textbase) rather than at deep levels (situation model), tasks that make the goal state explicit, or texts that make the situation model explicit do not reflect the full range of the levels of comprehension processes that are possible in the construction-integration model. For example, standardized tests of reading comprehension, typically only require the reading of short passages, and the answering of multiple choice questions (RAND, 2002), thereby restricting the kinds of comprehension that are assessed (Pearson & Hamm, 2005). On the other hand, longer passages and tasks that require the building of a situation model may be necessary to assess the full range of (deep or surface) levels of comprehension.

To assess deeper learning Kintsch (1998; Kintsch & Kozminsky, 1977; Kintsch, & van Dijk, 1978) has advanced the use of summary writing. Summary writing is the process of
determining what content in a passage is most important and transforming it into a succinct statement using one’s own words (Brown & Day, 1983; Kintsch & Kozminska, 1977; Stein & Kirby, 1992; Winograd, 1984). An important feature of using this measurement method to assess comprehension is that it relies on the scoring of free response data. In this respect, it can be classified as a generative or expressive measure of comprehension. Generative tests can be contrasted with receptive tests of knowledge, which require participants to select an answer from among a list of options, rather than to retrieve or construct the answer from memory (Schmidt & Bjork, 1992). However, because effective summary writing requires readers to use judgment and effort to intelligently condense and structure what they will retain as the gist of the text, it can be distinguished from memory recall in which all of the information available in memory is reproduced (Kintsch, & van Dijk, 1978).

Summarizing a text can be considered to be a form of meaning construction. In attempting to restate the content in a more concise form, summary writers must use their own words, balancing the need to be complete with being brief. Competent summary writing requires that interpretive inferences are made that involve elaborations, reordering, and generalizations (Kintsch, & van Dijk, 1978). In so doing, summarizing a text requires the summarizer to work with the meaning of the text at a deep level: summary writers must compare the text meaning with how they have conveyed it in the summary, gradually making the new knowledge their own. Writing a summary of a text may be most effective in assessing deeper levels of comprehension when it is done in the absence of the text. Having the original text present during the writing of the summary may encourage a copy and delete strategy, and, consequently, less active processing than required when a writer has to produce a summary after the target material has been removed (Hidi & Anderson, 1986; Kirby & Pedwell, 1991).
In their studies of planning skills for summary writing among students of different age groups (Grades 5, 7, 11 and college students), Brown, Day, and Jones (1983) indicated that “the ability to work recursively on information to render it as succinctly as possible requires judgment and effort, knowledge, and strategies ….. As such, the ability to provide an adequate written summary of a lengthy text is a late developing skill that continues to be refined throughout the school years” (p. 977). Summary writing, as a method of assessing reading comprehension performance, is included in Study Two.

**Summary and Limitations**

Presently, Kintsch’s model is considered to provide the most comprehensive account of the processes and products of comprehension (see McNamara & Magliano, 2009, for a review of influential theories of comprehension). However, several limitations to the model can be noted. First, the construction-integration model is intended to be a model of adult skilled reading. As noted previously, proficiency at the lowest level of surface processing, that is, in word reading ability, is assumed. This assumption may apply to the word reading of skilled adults, but is less likely to be applicable in younger or less skilled readers. Therefore, in a population where proficiency in this skill cannot be assumed, it is important that its influence be considered. Second, by outlining three interacting levels of comprehension (surface, texbase, and situational model) a useful theoretical tool is provided to assess at what level reading comprehension difficulties may occur. However, this requires that reading comprehension tasks be designed in a manner that appropriately targets for assessment each of the three levels of comprehension. Third, although Kintsch’s model highlights the importance of readers engaging in reasoning and in active and effortful processes to the achievement of deeper levels of comprehension, elaboration of these processes is not provided in his model.

**Summary of Models and Theories**

24
The overarching purpose in reviewing the models and theories above was to present what each of them offers in their explanations of what effective reading comprehension depends upon. By and large, these models are not mutually exclusive, but rather focus on different aspects of comprehension. Unfortunately, there does not exist presently a comprehensive model of reading comprehension.

The simple view of reading indicates that reading comprehension skill may not develop adequately because of weakness in the decoding component, the linguistic component, or both (Gough & Tunmer, 1986; Hoover & Gough, 1990). The lexical quality hypothesis (Perfetti, 2007; Perfetti & Hart, 2001) stresses the importance in effective reading of lexical quality, defined as accessible knowledge of a word’s form (phonology and orthography) and meaning. Accordingly, in any given text there is a risk of comprehension failing because the reader either lacks the necessary word representations, or the processes that connect word meanings to text representations lack sufficient high quality word representations with which to work. Perfetti and colleagues have argued that it is multiple and effective experiences with words in print that lead to the development of high quality word representations (Bolger et al., 2008; Frishkoff et al., 2008; Landi & Perfetti, 2006). Kintsch’s construction-integration model outlines three interacting levels of comprehension, (surface, textbase, and situation model), thereby providing researchers with a useful theoretical tool to assess at what level comprehension difficulties may occur. Clearly, understanding individual differences in reading comprehension requires that skill in both lower-level surface comprehension (word knowledge) and in higher levels (textbase and situation model) be considered.

Effective comprehension, however, also places demands on other cognitive processes and resources. First, a reader’s ability to hold, construct and integrate the ideas read in a text with what the reader already knows also places demands on other cognitive resources such as
working memory. Working memory resources are correlated with both children’s and adults’ text comprehension (Cain, Oakhill, & Bryant, 2004; Daneman & Merkle, 1996). The boundaries placed on comprehension by a limited working memory capacity are acknowledged in Perfetti’s verbal efficiency theory and Kintsch’s construction-integration model. Second, as outlined in Kintsch’s construction-integration model, effective comprehension requires the reader to engage in reasoning and to utilize conscious and in many cases intentional strategies. Both reasoning ability and the use of active strategies have been found to be related to reading comprehension performance (Brooke, Mulligan, Walton, 2005; Cain, 1999; Fuchs & Fuchs, 2005; Gersten, Fuchs, Williams, & Scott, 2001; Kozminsky & Kozminsky, 2001; Swanson, Kehler, & Jerman, 2010; VanderVeen, Huff, Gierl, McNamara, Louwerse, Graesser, 2007), and also should be considered when investigating possible factors underlying reading comprehension difficulties.

**Factors Contributing to Poor Comprehension**

According to the theories reviewed, effective reading comprehension may fail to develop for a variety of reasons. They include, but are not limited to, weaknesses in (a) accurate and efficient word reading skill and the antecedents that support its development (i.e., phonological awareness, orthographic knowledge, morphological knowledge, vocabulary, print experience), (b) the ability to construct a coherent and integrated representation of the meaning of the text, and (c) relevant cognitive strategies that support the effective construction of meaning. In addition, some individuals may have low general ability (i.e., working memory capacity, fluid intelligence) thereby limiting their skill in reading comprehension (Alloway & Alloway, 2010; Carretti, Borella, Cornoldi, & De Beni, 2009; Floyd, Bergeron, & Alfonso, 2006; Swanson, Zheng, & Jerman, 2009). Those whose skill is low in these various contributing factors will struggle with reading comprehension, and those whose skills are low on many of these factors will struggle more. Individuals who struggle with reading comprehension will enjoy reading less, creating a vicious cycle. To ensure that
reading comprehension difficulties are identified as early as possible, longitudinal studies that ascertain the extent to which early reading comprehension performance is predictive of future reading comprehension outcomes are necessary.

Because there are many factors that contribute to reading comprehension and because skill in these factors varies among readers, it is important that the specific sources of weaknesses of those who struggle with reading comprehension be identified. Two sub-groups of poor comprehenders warrant mention. The comprehension problems of the first sub-group, referred to as having a specific reading disability, are expected, given their word-level reading problems (Shaywitz & Shaywitz, 2005; Vellutino et al., 2004). Their reading problems are manifested in extreme difficulties in acquiring basic reading sub-skills such as word identification and phonological (letter-sound) decoding (see Vellutino et al., 2004, for a review). The comprehension problems of the second sub-group, referred to as having a specific comprehension deficit (Oakhill, Hartt, & Samols, 2005) are unexpected, given their age-appropriate word reading skills (see Cain & Oakhill, 2009, for a review). The unexpectedness of their comprehension problems, however, is not total as some language comprehension difficulties (e.g., weaknesses in inference making, comprehension monitoring, and integrating information) that interfere with reading comprehension have been observed (see Cain & Oakhill, 2009, for a review). In contrast to the vast amount of research that has been devoted to understanding the comprehension difficulties of children with specific reading disabilities (see Vellutino et al., 2004, for a review), less research has been devoted to understanding the nature of the comprehension difficulties of this second sub-group of poor comprehenders.

If we are to encourage the development of timely and appropriate instructional programs, longitudinal studies that examine the developmental course of reading comprehension outcomes, and studies that identify the locus of the comprehension
difficulties of poor comprehenders, are necessary. With these ends in mind, two studies are included in my dissertation. Study One examines the development of reading comprehension achievement over time to determine whether reading comprehension difficulties persist over time. To address the complex nature of the reading comprehension construct, Study Two uses the theoretical bases provided by the lexical quality hypothesis (Perfetti, 2007; Perfetti & Hart, 2001) and the construction-integration model (Kintsch, 1998) to examine the nature of the comprehension difficulties of adolescent poor comprehenders having a specific comprehension deficit.

**Details of Dissertation Studies**

**Study One: Development of Reading Comprehension**

The primary aim of Study One is to examine the development of reading comprehension achievement over time. Considerable evidence suggests that differences between children’s reading ability persist over time (Juel, 1988; Scarborough, 1998; Cunningham & Stanovich, 1997): children who are successful in the early grades tend to remain at the top of the achievement distribution in the years that follow, whereas children who have difficulty learning to read in the early grades tend to remain at the bottom. However, stability in the rank ordering of children’s reading achievement does not imply the absence of change in variability in achievement. Indeed, an important question is the extent to which individual differences in children’s reading achievement may increase or decrease over time. To account for each of these possibilities, two models have been proposed (Leppänen, Niemi, Aunola, & Nurmi, 2004; Parrila, Aunola, Leskinen, Nurmi, & Kirby, 2005; Phillips, Norris, Osmond, & Maynard, 2002; Stanovich, 1986). The cumulative model proposes that individual differences in reading achievement increase over time (Leppänen et al., 2004; Stanovich, 1986). According to this view, those with an initial advantage in reading, or in the prerequisite cognitive and linguistic skills to support its development, will improve their reading skill at a faster rate compared to those who begin the process with lower skill levels.
In contrast, the compensatory model proposes that children with initially lower skill may show faster growth relative to those with initially higher skill, perhaps as a result of extra instruction, thereby allowing them to narrow, and for some to close, the achievement gap over time (Parrila et al., 2005; Phillips et al., 2002). Such variability in relative reading achievement—that is, how individual differences in reading achievement may change over time—suggests that the development of reading comprehension may follow different pathways as outlined in the stability, cumulative, and compensatory models. Because the effect of measurement error has generally not been addressed in longitudinal studies in which change has been observed (Ferrer & Zhang, 2009), it is important that appropriate methods be used for distinguishing genuine behavioral change and apparent changes that are due to measurement errors.

Study One of my dissertation reports results of a longitudinal study of students that examined the prevalence of the stability, cumulative, and compensatory models in reading comprehension development. Although this study allowed me to determine the extent to which relative reading achievement remains stable over time, it is limited in that it does not assist us in understanding the sources of difficulties of those students who struggle with reading comprehension. This issue is addressed in Study Two of my dissertation.

**Study Two: Unexpected Poor Comprehenders**

The principal aim of Study Two was to examine both lower level word reading skills and higher level cognitive processes that have been related to comprehension in order to understand more about reading failure. This study identified a group of readers who possess adequate skill in decoding, but demonstrate specific weaknesses in comprehension. In the literature, this sub-group has been referred to as having a specific comprehension problem (Oakhill et al., 2005). The apparent age-appropriate word reading and decoding ability of this
A group of children, coupled with their relatively weaker reading comprehension, indicates that their performance can be considered to be unexpectedly poor.

A crucial question concerning this sub-group of poor comprehenders concerns the specificity of their comprehension difficulties. Different researchers have used different approaches to investigating the source of this group’s reading comprehension difficulties (e.g., Cain, 1999; Cain & Oakhill, 1999; Cain, Oakhill, & Lemmon, 2004, 2005; Cutting, Materek, Cole, Levine, & Mahone, 2009; Landi, 2010; Landi & Perfetti, 2006; Nation, Adams, Bowyer-Crane, & Snowling, 1999; Nation & Snowling, 1998a, 1998b, 2000; Oakhill, 1982; Oakhill, Yuill, & Parkin, 1986; Stothard & Hulme, 1992, 1995; Yuill & Oakhill, 1991; Yuill, Oakhill, & Parkin, 1989). For example, when decoding (pseudoword reading) and nonverbal ability have been controlled, poor comprehenders, relative to more skilled comprehenders, show weaknesses in semantic processing at the word level (i.e., vocabulary), but this is in addition to relative weaknesses in word reading accuracy and reading rate (e.g., Nation & Snowling, 1998a, 1999). When word reading and vocabulary have been controlled, poor comprehenders, relative to more skilled comprehenders, show weaknesses in the higher order comprehension skills of inferencing, text integration, and integrating text information with their background knowledge (see Cain & Oakhill, 2007, for a review). However, a distinction between comprehension difficulties occurring at the textbase and situation level of comprehension was not made in these studies. It is, therefore, still unclear at what semantic level their comprehension difficulties occur.

There is a need to clearly lay out the theoretical basis for understanding at what level this group’s comprehension difficulties occur. In doing so, I draw on Perfetti’s lexical quality hypothesis and Kintsch’s construction-integration model. As this literature review has outlined, there are several major levels, and many possible gradations within levels, in which comprehension difficulties may occur. As argued in the lexical quality hypothesis, at lower
levels, inefficient decoding processes and difficulty in accessing relevant meanings of words may contribute to comprehension difficulties. This may be in addition to word level weaknesses in morphological and orthographic knowledge. Further, previous research indicates that individual differences in reading experience are positively associated with word knowledge (e.g., Perfetti & Hart, 2001; 2002; Landi, 2010; Stanovich, 1986). As described in the construction-integration model, at the deepest level, poor comprehenders may have weaknesses at the situation model level of comprehension. This may be in addition to weaknesses at the textbase level of comprehension (micro- and macro-propositions). However, as outlined earlier, any relative weaknesses at the textbase and situation model level of comprehension may also be attributed to individual differences in working memory capacity or students’ approaches to learning (deep and surface).

In order to investigate whether comprehension difficulties of unexpectedly poor comprehenders at the textbase and situation model level are indeed unexpected, it is necessary to control for lower level word reading skills (word reading accuracy, reading rate, and phonological decoding) and general abilities (verbal and nonverbal). To my knowledge, the existing research on this group’s reading comprehension difficulties has not included all of these controls in one study. A systematic investigation of the effects of these controls is necessary to find out at what level their comprehension difficulties occur. Indeed, an important question that arises from previous research is whether the same individuals will be identified when different control procedures are used.

A second limitation of previous research is that generally only one contrast group has been used to investigate the possible sources of weaknesses in unexpectedly poor comprehenders (Cain, 2006; Cain & Oakhill, 2006; Cain, Oakhill, Barnes, & Bryant, 2001; Cain, Oakhill, & Elbro, 2003; Cain et al., 2004, 2005; Oakhill, 1982; Oakhill et al., 2005; Nation & Snowling, 2000, 1999, 1998a, 1998b; Palladino, Cornoldi, De Beni, & Pazzaglia,
rather than several groups who differ in their level of skill in reading comprehension. Differences in the overall mean comprehension skill level of the comparison group used (average versus above average) may lead to different findings for the variables of interest, and hence have implications for targeting specific areas for remediation.

A final limitation of previous research is that the majority of studies have been carried out with elementary school aged participants. Yet, developmental growth in word reading and comprehension skill continues as children proceed in their schooling (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996). At the high school level, students face increasingly complex narrative and expository texts, thereby placing progressively more demands on their comprehension skills. Because students’ performance on these more difficult texts is a strong indication not only of successful high school completion (Kamil, 2003; Snow & Biancarosa, 2003), but also of college-readiness and work-related reading tasks (ACT, Inc., 2006; Alliance for Excellent Education, 2005; STAND UP Coalition, 2006), it is important that further research be directed towards an older group of children.

Study Two investigates the sources of reading difficulties of unexpected poor comprehenders drawn from a sample of high school students. It systematically uses five control variables (word reading, pseudo-word reading, reading rate, non-verbal ability and verbal ability) and examines performance on a range of reading comprehension tasks, designed to reflect the distinction that Kintsch makes between textbase (that includes micro- and macro-propositions) and situation model comprehension. The study addresses whether (a) the comprehension weaknesses of unexpected poor comprehenders can be attributed to concurrent weaknesses in performance of control variables, that in previous studies, have not been explicitly controlled; (b) different control procedures lead to the identification of different groups of poor comprehenders; (c) higher level semantic weaknesses in textbase and
situation model comprehension remain when word reading, phonological decoding, reading rate, non-verbal ability, and verbal ability are controlled; (d) observed comprehension weaknesses at the textbase or situation model level can be attributed to relative weaknesses in performance for word related (i.e., phonological awareness, orthographic knowledge, and morphological knowledge), or individual differences variables (i.e., working memory, approaches to learning, and exposure to print); and, (e) any observed differences depend upon the nature of the contrast used (average versus above-average comprehenders). In addressing these questions, this study uses a multi-componential approach to address the complexity of the reading comprehension construct discussed in the literature review. Specifically, this study utilizes a design that recognizes that individual differences in reading comprehension require skill in lower-level surface comprehension (word knowledge), as outlined in the lexical quality hypothesis, and in higher levels (textbase and situation model), as per the construction-integration model. By including both a wider range of control procedures that have not been controlled in previous studies of poor comprehenders and a range of reading comprehension tasks at the textbase level and situation model levels, this study is intended to inform our theoretical understanding of the reading comprehension construct, and provide information that might be used to assist those adolescents who struggle with effective reading comprehension.
References


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


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


CHAPTER 2: A LONGITUDINAL STUDY OF READING COMPREHENSION ACHIEVEMENT FROM GRADES 3 TO 10: INVESTIGATING STABILITY, CUMULATIVE, AND COMPENSATORY MODELS

Abstract
This longitudinal study of 78 Canadian English speaking students examined the prevalence of pathways for stability, cumulative growth, and compensation in reading comprehension development. Archival government-mandated assessments of reading comprehension at Grades 3, 6, and 10, and the Canadian Test of Basic Skills measure of reading comprehension administered at Grade 10 were used. Probabilities of later-grade reading achievement categorization conditioned on earlier-grade reading achievement were computed, the prevalence of developmental paths was estimated, and tests of regression to the mean were conducted. Most changes in relative achievement were attributed to regression towards the mean. Overall results suggest considerable stability across time.

Introduction
In contrast to the vast amount of research that has been carried out over the last 40 years on the development of word reading skill, less research has been conducted investigating the development of reading comprehension (RAND Reading Study Group, 2002). Empirical research devoted to understanding the patterns of stability and change in reading comprehension performance is central to understanding reading development. Knowledge regarding different growth patterns for sub-groups of children assists us in ascertaining the extent to which early performance is predictive of future outcomes thereby encouraging the development of timely instructional and remedial programs.

Three main views have been proposed to describe the development of children’s reading comprehension performance. The first view argues that reading comprehension development shows a high degree of stability (e.g., Cunningham & Stanovich, 1997).
Children, on average, who are successful in the early grades tend to remain at the top of the achievement distribution in the years that follow, whereas children who have difficulty learning to read in the early grades tend to remain at the bottom. However, stability in rank ordering does not imply the absence of change in variability in achievement, and an important question is the extent to which individual differences may, over time, increase or decrease. To account for each of these possibilities, two other models have been proposed. The cumulative model, alternatively called a Matthew effect (Stanovich, 1986), a fan-spread pattern (Cook & Campbell, 1979), or a cumulative reading trajectory (Leppänen, Niemi, Aunola, & Nurmi, 2004), proposes that individual differences increase over time. According to this view, those with an initial advantage in reading, or in the prerequisite cognitive and linguistic skills to support its development, will improve their reading skill at a faster rate than those who begin the process with lower skill levels. In contrast, the compensatory model proposes that children with initially lower skill may show faster growth than those with initially higher skill, perhaps as a result of extra instruction, thereby allowing them to narrow, and for some to close, the achievement gap over time (Leppänen et al., 2004; Parrila, Aunola, Leskinen, Nurmi, & Kirby, 2005; Phillips, Norris, Osmond, & Maynard, 2002).

In this study, the extent to which the stability, cumulative, and compensatory views describe the development of reading comprehension is investigated. To do this, data on the achievement of 78 students whose reading comprehension performance was measured at three times (Grades 3, 6, and 10) is analyzed.

**Patterns of Reading Comprehension Growth**

**The Stability Model**

Considerable stability in reading achievement across time has been reported in several large samples, with relatively strong correlations (.59 to .88) between measures of reading taken one and six years apart (Badian, 1999; Phillips et al., 2002; Shaywitz et al., 1995). Correlation coefficients have generally been stronger when the time periods between testing...
intervals are shorter than when longer time periods are used, and when the same measure, rather than different measures of reading comprehension, are used (e.g., Phillips et al., 2002). But this has not always been the case (Aarnoutse, van Leeuwe, Voeten, & Oud, 2001; Carreker et al., 2007).

Support for the stability model comes from a longitudinal study (Grades 1 through 9) by Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher (1996) in which developmental changes in reading ability (composite score of word identification, pseudo-word reading, and reading comprehension) were investigated using yearly assessments with a sample of 403 children classified into three groups in Grade 3: reading disabled-discrepant (actual reading achievement score at Grade 3 was at least 1.5 standard errors below the score predicted from the child's IQ regardless of level of reading achievement), reading disabled-low achievers (children who did not show a 1.5 standard error difference between actual and IQ predicted achievement but who had reading cluster scores below the 25th percentile), and not reading impaired (reading achievement above 25th percentile and IQ and reading achievement difference less than 1.5 standard errors). Individual growth curves were modeled for each group. They reported that the groups did not differ, on average, in the duration of reading skill development - on average each group reached a plateau in reading achievement at approximately age 15. The differences in reading scores between the two disabled reading groups were not found to be statistically significant at any grade, but for each grade the average difference between the two disabled groups were significantly lower than that of the not reading impaired group, thus supporting the stability model of reading development. However, the composite nature of their reading outcome measure (word reading, pseudo-word reading and reading comprehension) can be criticized as not being sensitive to the issue of factorial invariance across time (Bast & Reitsma, 1998). Specifically, although word reading and reading comprehension are highly related, there is evidence that the factor
structure and development paths for each of these variables are distinguishable (Oakhill & Cain, 2007; Perfetti, Landi, & Oakhill, 2005). The authors addressed this issue by conducting separate analyses for each of the three sub-tests that made up the reading-cluster score. Results from these subsequent analyses resulted in identical conclusions to the analysis of the composite measure. A second issue concerns the level of reading skill of the not reading impaired group: examination of the reading-cluster scores indicated that this group’s achievement was above-average, thereby leaving the issue of whether significant group differences would be observed had the comparison group included average rather than above-average readers a moot point. Further, although individual growth curve analyses are useful when the interest rests with the nature of change in reading skills for different groups over time, the overall finding of stability in relative reading achievement results reported by Francis et al. represent group averages and only for those subjects for whom the model was appropriate. Although the authors reported that few cases (less than 10%) were dropped because of the model being potentially inappropriate, questions regarding the developmental course of relative reading achievement for these individuals remains unresolved.

Several studies have found substantial stability in children’s reading ability groupings (e.g., Juel, 1988; Smith, 1997). For example, Juel followed a large minority, low socioeconomic status sample from Grades 1 through 4 and reported a .88 probability that a child would be a poor reader (defined as the bottom quartile) at the end of Grade 4 if the child had been a poor reader at the end of Grade 1. Conversely, she reported a .12 probability that a child would be at least an average reader (defined as the top three quartiles) at the end of Grade 4 if the child was a poor reader in Grade 1. She also reported a .13 probability that a child would be a poor reader at the end of Grade 4 if the child were an average reader in Grade 1.
In contrast, in a longitudinal study (Grades 1 to 6) of reading achievement in an unselected sample of children, Phillips et al. (2002) found evidence for the compensatory view, and argued for a reconsideration of the stability assumption. Specifically, they reported that poor readers (defined as those performing one standard deviation or more below the mean on a reading comprehension measure) in Grade 1 had a .53 probability of becoming average readers in Grade 6 (i.e., those performing within one standard deviation of the mean on a reading comprehension measure). Further, Phillips et al. reported that above average readers (those performing more than one standard deviation above the mean) in Grade 1 had a .47 probability of being average readers in Grade 6. Indeed, their year-to-year correlations of .72 to .77 between Grade 1 and Grade 6, and a Grade 1 to 6 correlation of .59 indicate that even relatively high year-to-year correlations do not necessarily translate into completely stable category membership when a long enough period of time is examined (see also Badian, 1999). Given that no test of achievement is entirely reliable, differences are to be expected. Of particular concern in the assessment of change are statistical effects associated with regression toward the mean. Regression toward the mean will seem to support the compensatory model if only or mainly weak readers are used. Further, when achievement groups are established, a continuous distribution of scores must be divided at arbitrary cutoff points. Due to the lack of perfect reliability, there is a good chance that children who are close to one side of the cutoff point on one occasion will obtain scores on the other side on another occasion (e.g., Shaywitz, Fletcher, Holahan, & Shaywitz, 1992). However, changes in relative performance may not arise simply from the unreliability of scores. For example, Phillips et al. (2002) found that a decline in relative achievement for their above-average group and an increase in their below-average group could not be accounted for entirely by regression toward the mean.

The Cumulative Model
At the heart of the Matthew effect is the assumption of reciprocal relationships between initial individual differences in skills (e.g., phonological awareness, word reading, vocabulary) and the cognitive, behavioral, and motivational consequences that arise from such initial differences (Stanovich, 1986). For example, children who are reading well and who have good vocabularies will read more, learn more word meanings, and hence read even better. In contrast, children with inadequate vocabularies, who read slowly and without enjoyment, read less and will be caught in a downward spiral of slower development of vocabulary knowledge inhibiting further growth in reading ability. Given that individual differences in reading achievement levels among students in Grades 8 to 12 are larger than the differences in Grade 1 (Daneman, 1996; Williamson, Applebaum, & Epanchin, 1991), this model is intuitively appealing. However, in a number of longitudinal studies investigating evidence for the cumulative model in reading achievement, there have been inconsistent results.

Evidence supporting a Matthew effect in reading comprehension achievement has been provided in several studies. For example, Carreker et al. (2007) reported that students who had relatively higher reading comprehension scores at Grade 3 had significantly greater reading comprehension growth in Grades 4 and 5 than those who had lower reading comprehension scores at Grade 3. Further, but qualified evidence, was provided by Vauras, Kinnunen, and Kuusela (1994) in a longitudinal study with children age 9 to 12. Reading comprehension was measured by asking the children to read, and then to write a composition from memory on the theme corresponding to the title of a text. The authors conducted a qualitative analysis of the responses and reported children’s achievement in a number of achievement categories (e.g., local vs. global understanding of the text). After 2 years, initially low-achieving children showed relative decline on most, but not all, of these achievement categories. They also reported that the cumulative effect was most pronounced
when the growth of children with initially low achievement was compared to that of children with initially high achievement, rather than to that of children with initially average achievement. Such findings suggest that the specific properties of the reading comprehension construct that are tapped by a particular measure and the comparison groups used may lead to different results across studies.

Bast and Reitsma (1998) suggested that support for the cumulative model requires evidence of stability in rank order and increasing individual differences. They qualified this second aspect with what they termed, within-construct developmental limits, the provision that a particular pattern may only be apparent during certain periods in reading comprehension development, after which individual differences may remain relatively stable for a period of time. Aunola, Leskinen, Onatsu-Arvilommi, and Nurmi (2002) argued that the cumulative model also implies an amplifying developmental process. This means that the earlier levels of the construct should influence the rate of increase or decrease of the same construct later on. Specifically, subsequent growth should be greater for those with higher initial levels than for those with lower initial levels. Accordingly, one would expect a positive relationship between initial level and subsequent growth.

Some evidence supports an increase in variance that is limited to the initial phase of reading acquisition (e.g., Bast & Reitsma, 1998; Parrila et al., 2005). In a longitudinal study (Grades 1 to 3) of Dutch children, Bast and Reitsma reported an increase in reading comprehension variance from the beginning of Grade 1 to the beginning of Grade 2, followed by a decrease to the end of Grade 2, and no change between the end of Grade 2 and the end of Grade 3. Findings of low stability in individual differences in reading performance led them to conclude a null result for the Matthew effect over the Grades 1 to Grade 3 period. Parrila et al. also reported increases in reading comprehension variance from Grades 1 to 2 in a longitudinal study (Grades 1 to 5) of English speaking children. Unlike Bast and Reitsma,
they tested the statistical significance of this increase and found it to be nonsignificant. They also reported significant decreases in variance between Grades 2 and 5. Further, contrary to the expectation of the cumulative model, Parrila et al. reported a significant negative relationship between initial level and subsequent growth: The higher the initial level of comprehension, the less improvement there was in it.

Overall, these studies provide considerable evidence against a Matthew effect in reading comprehension past the early elementary school years. However, this does not rule out the possibility that some children will show a Matthew effect, or that the negative effects associated with the Matthew effect may be limited to only low-starting children.
A special case of the cumulative model: The Fourth Grade slump. In contrast to the relatively rapid acceleration in reading achievement observed in the early elementary grades (Grades 1 to 3), studies have reported a slowing in growth of children’s reading achievement as they progress into the middle school years (Grades 4 and above) (Ferrer et al., 2007; Phillips et al., 2002). Although almost all children show improvement in their absolute level of performance as they progress through school, the rate of growth of some children may slow to such a level that they regress relative to other children. That some children demonstrate initially adequate reading achievement, but later experience difficulties has been referred to as the fourth grade slump (Chall, 1996), the time period in which this pattern has been most commonly observed, and can be considered a special case of the cumulative model. Leach, Scarborough, and Rescorla (2003) have referred to this profile as late emerging reading disabilities. In studies investigating this pattern of reading development, the prevalence of this sub-group of children has ranged from 6% to 20% of the respective samples (Badian, 1999; Cain & Oakhill, 2007; Juel, 1988; Leach et al., 2003; Kavale & Reese, 1992; Phillips et al., 2002).

A number of reasons have been offered for a slowing of growth. First, the fourth grade is considered the period in formal schooling when children make the transition between learning to read and reading to learn (Chall, 1996). At this time, heavier demands are made not just for deeper understanding of text but also for dealing with more challenging aspects of lower-level processes. Materials at the Grade 4 level and higher are different than those experienced in the primary grades (e.g., exposition vs. narration) placing more demands on children’s knowledge of specific text structures (Best, Floyd, & McNamara, 2008). Further, the unfamiliar vocabulary in the new texts may place increased demands on the decoding and oral language abilities of children as they proceed through the upper middle grades into high school where the importance of vocabulary and prior knowledge becomes greater (Best et al.,
2008; Nakamoto, Lindsey, & Manis, 2007). There is some evidence that the problem can be traced back to earlier weaknesses (e.g., Catts, Hogan, & Adlof, 2005; Spira, Bracken, & Fischel, 2005) thereby supporting the assumption of the Matthew effect for reciprocal relationships and the cumulative consequences for performance (Stanovich, 1986).

A good portion of the research concerned with this phenomenon has focused only on those children whose initial achievement is typical, and then falls to a level determined to indicate poor achievement. Aside from the difficulties associated with different methods used across studies to establish cut-off scores for poor achievement, such an approach neglects the full range of students who may experience a relative decline in their reading performance. Specifically, it fails to consider that some children who start out with initially adequate skills may experience a relative decline, but the decline may not be sufficient for them to have dropped to a level categorized as poor achievement.

The Compensatory Model

The compensatory model proposes that children with initially lower skill may show faster growth than those with initially higher skill thereby allowing them to reduce, and, for some, to close the achievement gap over time (e.g., Aarnoutse et al., 2001; Aunola et al., 2002; Parrila et al., 2005; Phillips et al., 2002; Scarborough, 1998). Several explanations have been offered for the observed differences in the degree of compensation exhibited by slow starting individuals. A higher degree of compensation has been observed in those children with higher cognitive and linguistic abilities, who come from homes with higher socio-economic status, demonstrate positive instructional and motivational effects, and whose reading strategy use is more prevalent (Bruck, 1990; Leach et al, 2003; McCall, Hauser, Cronin, Kingsbury, & Houser, 2006; Shaywitz et al., 2003).

A number of studies have reported that some children with low initial performance levels can improve sufficiently over time to perform within the average range on reading
comprehension measures (Bruck, 1992; Ghelani, Sidhu, Jain, & Tannock, 2004; Parrila, Georgiou, & Corkett, 2007; Phillips et al., 2002; Shaywitz et al., 2003). These studies have been predominantly carried out with children with learning disabilities. However, there is also some suggestion in the literature that, on average, these children rarely perform above the low average range (Shaywitz et al., 2003; Scarborough, 1998).

In sum, that some initially low-starting children may demonstrate compensation does not necessarily imply that the compensation they exhibit will be sufficient for them to close the gap entirely with their higher-starting peers (Aarnoutse et al., 2001; Parrila et al., 2005; Shaywitz et al., 1995). Such findings lend support to the notion that the development of reading achievement with some low-starting children might be best represented by a pattern of compensation that accommodates stability (Parrila et al., 2005). However, it should be noted that the pattern of improvement for low-starting children is consistent with the statistical artifact of regression toward the mean (but see Phillips et al., 2002).

**Summary of Findings and Limitations of Existing Studies**

Overall, the review of the literature suggests the following: One, the pattern of increasing individual differences in reading between initially low- and high-starting children may be limited to the first few years of formal schooling; Two, a slowing of growth in reading achievement generally occurs in most children around Grade 4; And, three, this slowing of growth in achievement is generally greater for initially higher starting children than for lower starting children, with the observation that the growth in achievement of some strong-starters may slow down to the point that they become identifiable as having reading problems. These observed patterns are reflected in the cumulative model. Conversely, some initially low-starting children may demonstrate sufficient growth for them to perform within the average range in later years as reflected in the compensatory model. Taken together, these findings suggest that the development of reading comprehension may follow multiple
pathways: the stability model proposes no changes in relative achievement over time, the cumulative model proposes decreases, and the compensatory model increases for some individuals. A stringent test of the prevalence of any of these three pathways would require that there be more than two testing occasions. Further, as suggested by Bast and Reitsma’s (1998) notion of within-construct developmental limits, the existing conceptualizations do not accommodate the possibility of variability in pathways. This variability may be expressed in a manner that is not limited to the expectations set out by any one particular model. For example, as described in the cumulative model, individuals may experience a decrease in their reading achievement during one period of time; however, the observed decrease may be developmentally limited. It may be followed by either an increase in achievement, as described in the compensatory model, or by a period of stability. Alternatively, the achievement of some individuals may be stable during one period of time, and be followed by a period in which their achievement increases or decreases.

The vast majority of longitudinal studies investigating individual differences in the development of reading comprehension have limited their investigation to the early and middle school years (Grades 1 to 6). However, as many authors have noted, skill in reading comprehension continues to develop past the middle school years (e.g., Bast & Reitsma, 1998; Francis et al., 1996; Parrila et al. 2005; Phillips et al., 2002; Shaywitz et al., 1999). Further research is required to investigate the nature and prevalence of the developmental paths in students’ reading comprehension achievement as they progress into high school.

Previous studies investigating individual differences have commonly assigned students to various achievement groupings (e.g., below-average/poor, average/good) based on the rank ordering of individuals on a particular task in relation to the performance of other children included in that particular study. In this manner, performance groupings refer only to the rank ordering of individuals within a given sample. The recent use of standards-based
assessments by educational policy makers attempts to overcome this limitation by establishing criterion-referenced levels of performance (Wixson & Carlisle, 2005).
Specifically, designation of a student’s performance as average or below-average refers not to the student’s relative position in a given sample’s distribution, but rather to how well the student has met some pre-established criteria. This study addresses this issue by assigning students to achievement groupings using criterion-based scores from government reading comprehension tests.

Present Study
This study was a retrospective longitudinal investigation of the reading comprehension achievement of students at three time points in their schooling (Grades 3, 6, and 10). My primary concern is to investigate the degree to which the development of reading comprehension is represented by patterns set out in the stability, cumulative, and compensatory models, and whether the observed patterns depend upon initial levels of reading achievement. Specifically, the study addresses the following questions: (1) What are the relationships among the four comprehension measures (one in each of Grades 3 and 6, two in Grade 10)? (2) What are the probabilities of later reading achievement categorization given earlier categorization? and (3) What are the developmental paths of reading achievement for Grades 6 and 10 based on Grade 3 achievement, and what is the prevalence of these paths at Grade 10?

Local Context
In Ontario, large-scale mandated assessments of reading are administered to all elementary students in Grades 3 and 6 (ages 8 and 11 respectively). Relative reading achievement in these grades is reported using four main performance levels, (details provided below). At the high school level the Ontario Secondary School Literacy Test (OSSLT), which
includes reading and writing components, is administered in Grade 10 (age 15). Students are required to achieve a passing mark on the OSSLT to graduate from high school. These standard-based tests, designed and administered by The Education Quality and Accountability Office (EQAO), are intended to measure student achievement against curriculum expectations. Results of these tests yield individual, school, school board, and provincial data on student achievement that are intended to help guide improvement planning. EQAO, however, did not maintain records in a manner that would enable large-scale longitudinal analyses of student achievement at the individual level. Access was limited to files from which names, Ontario education numbers and other personal identifiers had been removed, and files were not linked across grades. Requests to access individual student records could be made to school boards, but required student and parental consent.

As the interest in this study concerns the relative achievement of individuals over time, a retrospective longitudinal design is utilized. EQAO assessment data on reading achievement (Grades 3, 6 and 10) were used to track the development of reading comprehension achievement in this cohort. A commercially available reading comprehension test administered in Grade 10 was used to create a broader measure of Grade 10 reading achievement.

Method

Participants

The participants were 78 Grade 10 students whose first language was English. The mean age of the participants was 15 years 4 months ($SD = 7.5$ months); there were 48 girls and 30 boys. They were enrolled in a mid-sized rural Ontario publicly funded Catholic high

2 Students who have been unsuccessful twice on the OSSLT may enroll in a specially designed literacy course. Successful completion of this course allows students to obtain their high school diploma.
school that serves approximately 1000 students from surrounding communities, representing a range of socio-economic backgrounds. A comparison of demographic characteristics of the district served by the high school to that of the province (Statistics Canada, 2001) indicated that it had a greater reliance upon manufacturing and construction (34% vs. 22%), a greater portion with less than a high school diploma (28% vs. 20%), and a smaller proportion with a university education (11% vs. 24%). Aside from those people with an aboriginal identity (9% vs. 2%), there are very few visible minorities (1% vs. 19% provincially). English is the first language learned by a large segment of the population (88% vs. 71% provincially). In the present sample, 18% of the students (N = 14) had been identified with special learning needs; half of these students had been formally identified (7 learning disabled, 1 gifted, 1 mild intellectual disability, 1 behavioral), and the remaining students had been informally identified with only an individual educational plan (1 mild hearing loss; 3 unspecified). Students who participated in the study were enrolled in either the fall or winter sessions of Grade 10 English academic or applied streams of study (September to January N = 56; February to June, N = 22). Whereas the academic stream prepares students for university, the applied stream prepares students for vocational college, or for the workplace. Student and parental permissions were obtained.

Measures

**Demographic questionnaire.** Participants completed questions regarding their gender, level of academic study, first language learned, and age at the start of the study.

**Reading comprehension measures.** Four measures of reading comprehension measures were used.

**Grades 3 and 6.** Archival data for the EQAO group-administered assessments of Reading, Primary Division (Grades 1 to 3) and Junior Division (Grades 4 to 6) in the Spring of students’ enrollment in Grades 3, and 6 were obtained. In the two assessment times in
which the present sample participated, 1999/2000 Grade 3 and 2002/2003 Grade 6, two separate reading passages were included and students responded to both open-ended and multiple-choice questions. In Grade 3, there were 18 open-ended and 20 multiple-choice questions (EQAO, 2000a). In Grade 6, there were 19 open-ended and 21 multiple-choice questions (EQAO, 2003a). At both grades, students’ reading achievement was reported using four provincial levels (PL1 through PL4) with PL3 set as the standard. EQAO also reported the following designations: Not Enough Evidence, for those students who had not demonstrated enough evidence of knowledge and understanding to be assigned Level 1; Insufficient Information to Score, for student work that was deemed to be insufficient to be given a level of achievement (e.g., if large sections of work were missing due to absence or another reason); Exempt and No Data, for students who were formally exempted from participation (e.g., if the test had to have been read to the student by a teacher or other adult); and Non-exempt students from whom EQAO did not receive completed assessment booklets (EQAO, 2000b; EQAO, 2003b).

The EQAO Grade 3 and 6 assessments took place over a three-week period in May of 2000 and 2003 respectively with five days (approximately 2.5 hours each day) devoted to administration of the assessments of reading, writing, and mathematics. In both grades, students took part in introductory activities with their classmates and then worked independently to read the passages, write their responses, and answer multiple-choice questions. Reliability data obtained from EQAO for the Grade 3 and Grade 6 reading assessment indicated an inter-rater reliability for the open-ended responses of approximately

---

3 PL 4: The student has demonstrated the required knowledge and skills. Achievement exceeds the provincial standard. PL3: The student has demonstrated most of the required knowledge and skills. Achievement meets the provincial standard. PL2: The student has demonstrated some of the required knowledge and skills. Achievement approaches the provincial standard. PL1: The student has demonstrated some of the required knowledge and skills in limited ways. Achievement falls much below the provincial standard.
60% based on the remarking of approximately 15% of student responses (M. Kozlow, EQAO, personal communication, February 11, 2009). For Grade 3, no reliability estimates were calculated; for Grade 6 Cronbach’s alpha for the multiple choice test and the constructed response items was .91 based on approximately 25% of students (Kozlow).

**Grade 10 Ontario Secondary School Test.** Archival data for EQAO’s Ontario Secondary School Literacy Test (OSSLT) written in the spring of students’ enrollment in Grade 10 were obtained. The OSSLT assesses both reading comprehension and writing skills. The OSSLT was administered in two 75-minute sessions in specially assigned classrooms on the day of testing by school staff. Although results for reading and writing are normally not reported separately, and students’ overall achievement is normally reported using a dichotomous scale (successful/unsuccessful), EQAO provided a data file containing scores for each of the questions for the present study; the raw scores for only the reading component were used here.

The 2007 reading component of the OSSLT consisted of five reading selections (information paragraph, news report, dialogue, narrative, and graphic text) that varied in length from a single paragraph to two pages. Students responded to 31 multiple-choice and 4 open-response questions that assessed their understanding of explicitly stated ideas and information, implicitly stated ideas and information, and connections made between information and ideas in a reading selection with students’ knowledge and experience (EQAO, 2007). The raw scores assigned to each reading comprehension item (1 mark for each multiple-choice item; maximum 3 marks for each open-ended response) were used in the analyses that follow (total maximum score 43). EQAO reported an alpha reliability coefficient of 0.87 for the combined reading and writing items (M. Kozlow, EQAO, personal communication, January 30, 2008). In this sample, the Cronbach’s alpha for the combined reading and writing items was .76 and .74 for the reading items.
Canadian Test of Basic Skills (Nelson Education, 1998). Test-R (Reading) of the Canadian Tests of Basic Skills (CTBS) – English (Form K, Level 16) is a sub-test of the CTBS norm-referenced achievement battery. Form K of the CTBS is adapted from Form K of the Iowa Test of Basic Skills. The test assesses the factual and inferential understanding of reading passages. Following the instructions outlined in the publisher’s manual, this test was group administered with 40 minutes to answer 44 questions based on five reading passages (two narratives, one poem, and two expositions). Multiple-choice questions follow each passage with five answer choices per item. Each correct response scores 1 point for a maximum of 44. I administered the CTBS reading comprehension test. The Cronbach’s alpha value of .83 in this sample is similar to the KR-20 value of .89 reported in the manual.

Results

Preliminary Analyses

In the first phase of analysis, the reading achievement and demographic characteristics of the students included in the present sample were compared to those reported at the provincial and school levels. Distributions into the EQAO reading achievement categories for Grade 3 (2000) and 6 (2003) and representation based on gender, level of academic study, and success rate for the OSSLT (reading and writing) were examined. Overall, the percentages of students whose achievement fell into each of the provincial levels for Grade 3 and 6 in the present sample were similar to those reported at the provincial level (Not Enough Evidence and Insufficient Information to Score, 11.5%, 7.7% versus 13%, 10%; PL1 5%, 7.7% versus 8%, 7%; PL3 36%, 32.1% versus 30%, 27%; PL4 6.4%, 5.1% versus 6%, 9% respectively). For the OSSLT, there was a greater representation of girls than boys in the present sample (61% vs. 39%), higher than reported at the provincial level (49% vs. 51%), but similar to the representation at the school level (57% vs. 43%). The majority of students (73%) were enrolled in the academic stream of study, consistent with the provincial representation, but higher than reported at the school level (63%). Students in the present
sample had a higher OSSLT pass rate than reported at the provincial or school levels (89.7% vs. 84% and 86% respectively).

**Coding the Grade 3 and 6 Data**

Prior to carrying out the analysis necessary to address the study’s questions, the EQAO data were coded. To maintain the four point scoring method (PL 1 through 4) used by EQAO in the Grade 3 and 6 assessments, students whom EQAO had coded as Not Enough Evidence, Insufficient Information to Score, Exempt and No Data, were re-coded as PL1 (Grade 3, N = 9; Grade 6, N = 6).

**Establishing the Grade 10 Composite and Achievement Groups**

To create a broader measure of Grade 10 reading achievement, a composite score was first calculated by averaging z-scores for the two Grade 10 tests. To form Grade 10 achievement groups that were proportional to those in Grades 3 and 6 (i.e., had approximately the same percentages of students in the four levels), the percentages of students in each of the Grade 3 and Grade 6 achievement groupings were averaged. Then the Grade 10 z-score was broken into groups with approximately the same proportions. Students whose reading achievement was classified as poor in Grade 10 represented 17.5% of the sample (Grade 10 composite z-score range -2.14 to -0.92), below-average (31.2%, range -0.90 to 0.09), average (44.6%, range 0.12 to 1.19), and above-average (6.7%, range 1.25 to 1.65).

**Descriptive Statistics**

Table 2.1 reports descriptive statistics for the four reading comprehension measures. Based on the norms reported in the CTBS manual, the mean grade equivalent for the Grade 10 CTBS was in the range expected for Grade 10 (M = Grade 10.2, SD = 3.7). The OSSLT reading measure was negatively skewed. Transformations of this variable improved its distribution but did not change its correlations with other variables. Therefore, all subsequent analyses were performed on the raw scores.
Relationships among Reading Comprehension Measures

Table 2.2 presents the correlations between the four reading comprehension measures. These variables were significantly correlated ($p < .001$) in the moderate range (range 0.54 to 0.65). The correlations of the Grade 10 reading comprehension composite and Grade 10 achievement groups with each of the other variables are also shown in Table 2.2. Both the Grade 10 composite and achievement groups were more highly correlated with the Grade 3 (.69 and .67 respectively) than with the Grade 6 score (.64 and .63 respectively), but all correlations are substantial.
Table 2.1

*Descriptive Statistics (N = 78)*

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>M</th>
<th>SD</th>
<th>Skew</th>
<th>SE</th>
<th>Kurtosis</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 3</td>
<td>1</td>
<td>4</td>
<td>2.37</td>
<td>0.84</td>
<td>-0.12</td>
<td>0.27</td>
<td>-0.67</td>
<td>0.54</td>
</tr>
<tr>
<td>Grade 6</td>
<td>1</td>
<td>4</td>
<td>2.42</td>
<td>0.81</td>
<td>-0.34</td>
<td>0.27</td>
<td>-0.60</td>
<td>0.54</td>
</tr>
<tr>
<td>Grade 10 OSSLT</td>
<td>18</td>
<td>42</td>
<td>35.97</td>
<td>4.63</td>
<td>-1.25</td>
<td>0.27</td>
<td>1.99</td>
<td>0.54</td>
</tr>
<tr>
<td>Grade 10 CTBS</td>
<td>7</td>
<td>41</td>
<td>26.42</td>
<td>8.21</td>
<td>-0.21</td>
<td>0.27</td>
<td>-0.81</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Table 2.2

*Correlations between Reading Comprehension Measures*

<table>
<thead>
<tr>
<th></th>
<th>Grade 3</th>
<th>Grade 6</th>
<th>Grade 10 OSSLT</th>
<th>Grade 10 CTBS</th>
<th>Grade 10 Composite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 6</td>
<td>.55**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 10 OSSLT</td>
<td>.58**</td>
<td>.54**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 10 CTBS</td>
<td>.65**</td>
<td>.60**</td>
<td>.60**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Grade 10 Composite</td>
<td>.69**</td>
<td>.64**</td>
<td>.89**</td>
<td>.89**</td>
<td>1</td>
</tr>
<tr>
<td>Grade 10 Groups</td>
<td>.67**</td>
<td>.63**</td>
<td>.80**</td>
<td>.86**</td>
<td>.93**</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.001 level (1-tailed).
Relationship between Initial Levels of Reading Comprehension Achievement and Changes in Reading Comprehension Achievement

To investigate the relationship between initial levels and changes in reading comprehension performance, three sets of conditional probabilities were computed using the crosstab function of the statistical program SPSS (version 17): (a) Grade 6 reading achievement categorization given Grade 3 reading achievement categorization (see Table 2.3), (b) Grade 10 categorization given Grade 6 categorization (Table 2.4), and (c) Grade 10 categorization given Grade 3 categorization (Table 2.5). The four EQAO PLs (PL 1-4) were used to categorize students’ Grade 3 and 6 reading achievement. For Grade 10, students’ reading achievement was categorized using the four reading achievement categories (poor, below-average, average, above-average outlined above).

With some exceptions, the results for the three analyses are similar (see Tables 2.3, 2.4, and 2.5). Overall, a large degree of stability was observed. Students at PL1 and PL3 were most likely to maintain the same level of achievement in the later grade as in the earlier grade (range .54 to .69). There was a general trend for all groups that the probabilities of maintaining the same level of achievement would be stronger when later and longer periods of time were examined. In contrast, PL4 students showed low stability from Grade 3 to 6 or Grade 3 to 10 (range .20 to .40), but much greater stability from Grade 6 to 10 (.75).

For the weakest starters (PL1 and PL2), the probability of a relative increase in reading achievement was stronger in the analyses that examined the shorter time periods (Grade 3 to 6, and Grade 6 to 10), than the longer (Grade 3 to 10) time period (probability range .32 to .47 vs. .29 respectively). However, overall for each time period examined, the probabilities of these students maintaining their achievement categorization were strong (range .43 to .69). There was a low probability for the performance by PL2 students to decrease (range .11 to .20); students starting at PL1 could go no lower.
Table 2.3

*Conditional Probabilities of Reading Achievement Categorization in Grade 6 Conditioned on Grade 3 Achievement*

<table>
<thead>
<tr>
<th>Grade 3 Category</th>
<th>%</th>
<th>n</th>
<th>PL1</th>
<th>PL2</th>
<th>PL3</th>
<th>PL4</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL1</td>
<td>16.7</td>
<td>13</td>
<td>.54</td>
<td>.39</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>PL2</td>
<td>35.9</td>
<td>28</td>
<td>.11</td>
<td>.43</td>
<td>.46</td>
<td></td>
</tr>
<tr>
<td>PL3</td>
<td>41.0</td>
<td>32</td>
<td>.06</td>
<td>.25</td>
<td>.59</td>
<td>.09</td>
</tr>
<tr>
<td>PL4</td>
<td>6.4</td>
<td>5</td>
<td></td>
<td></td>
<td>.80</td>
<td>.20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>78</td>
<td>12</td>
<td>25</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td></td>
<td>15.4</td>
<td>32.1</td>
<td>47.4</td>
<td>5.1</td>
</tr>
</tbody>
</table>

*Note. Bold indicates probability that categorization remains the same.*
Table 2.4

Conditional Probabilities of Grade 10 Reading Achievement Categorization Conditioned on Grade 6 Achievement

<table>
<thead>
<tr>
<th>Grade 6 Category</th>
<th>%</th>
<th>n</th>
<th>Poor</th>
<th>Below-Average</th>
<th>Average</th>
<th>Above-Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL1</td>
<td>15.3</td>
<td>12</td>
<td>.58</td>
<td>.42</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PL2</td>
<td>32.1</td>
<td>25</td>
<td>.20</td>
<td>.44</td>
<td>.32</td>
<td>.04</td>
</tr>
<tr>
<td>PL3</td>
<td>47.5</td>
<td>37</td>
<td>.05</td>
<td>.22</td>
<td>.68</td>
<td>.05</td>
</tr>
<tr>
<td>PL4</td>
<td>5.1</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>.25</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>78</td>
<td>14</td>
<td>24</td>
<td>34</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>17.9</td>
<td>30.8</td>
<td>43.6</td>
<td>7.7</td>
<td></td>
</tr>
</tbody>
</table>

Note. Bold indicates probability that categorization remains the same.
<table>
<thead>
<tr>
<th>Grade 3 Category</th>
<th>%</th>
<th>n</th>
<th>Poor</th>
<th>Below-Average</th>
<th>Average</th>
<th>Above-Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL1</td>
<td>16.7</td>
<td>13</td>
<td>.69</td>
<td>.21</td>
<td>.08</td>
<td>-</td>
</tr>
<tr>
<td>PL2</td>
<td>35.9</td>
<td>28</td>
<td>.11</td>
<td>.61</td>
<td>.29</td>
<td>-</td>
</tr>
<tr>
<td>PL3</td>
<td>41.0</td>
<td>32</td>
<td>.06</td>
<td>.13</td>
<td>.69</td>
<td>.13</td>
</tr>
<tr>
<td>PL4</td>
<td>6.4</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>.60</td>
<td>.40</td>
</tr>
<tr>
<td>n</td>
<td>78</td>
<td>14</td>
<td>24</td>
<td>34</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>17.9</td>
<td>30.8</td>
<td>43.6</td>
<td>7.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. Bold indicates probability that categorization remains the same.*
Generally, the probabilities of adequately-starting students (PL3 and PL4) experiencing a relative decline were greater in the analyses that examined the shorter time periods (Grade 3 to 6, and Grade 6 to 10) than the longer time period (Grade 3 to 10) (combined probability of a one or two level decrease .27 and .31 respectively vs. .19). Further, although these probabilities were higher than the probability of relative increases in achievement (range .05 to 13), the probabilities for relative increases or decreases were both lower than the probabilities that relative achievement would remain the same (range .59 to .68). The highest probability of a relative decrease was observed for students starting at PL4 in the Grade 3 to 6 and Grade 3 to 10 periods (.80 and .60). However, given the low number of students in this group ($N = 4, 5$ respectively) caution in interpreting these results is suggested.

For the total sample, very few students (3.8%) experienced a two level change in achievement, with the same likelihood of this change occurring in any time period examined. A two level change was most prevalent for students with initial PL3 achievement. No child at PL4 experienced a two level drop.

In summary, consistent with the overall trend for the probabilities of maintaining the same level of achievement to increase when longer time periods were examined, a large percentage of the sample maintained relative stability in achievement (Grade 3 to 6, 50%; Grade 6 to 10, 59%; Grade 3 to 10, 64%).

**Developmental Paths**

The main interest in these analyses was to examine the developmental paths of students’ reading achievement across the three assessment times. Based on their level of reading achievement in Grade 3, the prevalence of the paths students’ reading achievement

---

4 There were four students with initial PL3 achievement who experienced a two category decline, two students with initial PL1 achievement, and one student with initial PL2 achievement who experienced an increase. One of these students experienced a two level change twice, first up and then down.
Developmental Paths of Grade 3 Achievement Groups. Table 2.6 reports the prevalence of the eight developmental patterns that were identified: (a) stability, (b) early increase, (c) late increase, (d) early decrease, (e) late decrease, (f) increase and decrease, (g) decrease and increase, (h) two category changes. The pattern of stability refers to students who maintained the same relative achievement level at each of the three testing occasions (38.5% total sample). The patterns of early increase (9%), late increase (9%), early decrease (9%), and late decrease (2.6%) refer to students who maintained stability in their relative achievement for two consecutive periods. The paths of increase/decrease (15.4%), and decrease/increase (9.0%) refer to students whose reading achievement category in Grade 10 was the same as in Grade 3, but different in Grade 6. The last path, two category change (7.7% of total sample), refers to students whose relative achievement changed by two levels across the three times. There was heterogeneity in the timing and direction of the six students who exhibited two level changes.\(^5\)

---

\(^5\) One student experienced a one level improvement after both Grade 3 and 6; two students experienced a two level drop in Grade 6 and maintained that level in Grade 10; one student demonstrated a two level decrease in Grade 10 after showing an increase in Grade 3 to 6; one student demonstrated a two level increase in Grade 6, but then dropped by two levels in Grade 10; one student showed a two level increase in Grade 10, after showing a one level achievement drop in Grades 3 to 6.
Table 2.6

*Prevalence (Percentages) of Development Paths in Grade 6 and 10 Based on Grade 3 Reading Achievement*

<table>
<thead>
<tr>
<th>Developmental Path</th>
<th>PL1 (n = 13)</th>
<th>PL2 (n = 28)</th>
<th>PL3 (n = 32)</th>
<th>PL4 (n = 5)</th>
<th>% Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stability</td>
<td>30.8</td>
<td>28.6</td>
<td>53.1</td>
<td>20.0</td>
<td>38.5</td>
</tr>
<tr>
<td>Early Increase</td>
<td>-</td>
<td>17.9</td>
<td>6.3</td>
<td>-</td>
<td>9.0</td>
</tr>
<tr>
<td>Late Increase</td>
<td>23.1</td>
<td>10.7</td>
<td>3.1</td>
<td>-</td>
<td>9.0</td>
</tr>
<tr>
<td>Early Decrease</td>
<td>-</td>
<td>3.6</td>
<td>9.4</td>
<td>60.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Late Decrease</td>
<td>-</td>
<td>3.6</td>
<td>3.1</td>
<td>-</td>
<td>2.6</td>
</tr>
<tr>
<td>Increase/Decrease</td>
<td>30.8</td>
<td>25.0</td>
<td>3.1</td>
<td>-</td>
<td>15.4</td>
</tr>
<tr>
<td>Decrease/Increase</td>
<td>-</td>
<td>7.1</td>
<td>12.5</td>
<td>20.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Two Category Changes</td>
<td>15.4</td>
<td>3.6</td>
<td>9.4</td>
<td>-</td>
<td>7.7</td>
</tr>
</tbody>
</table>

The prevalence of the development paths depended on students’ Grade 3 level of achievement. For students starting out at the highest level of achievement (PL4, N = 5), the pattern that best described their development was of *early decrease* (60%). In contrast, a pattern of *stability* (53%) best characterized the development of students with Grade 3 achievement at PL3 (N = 32). For students starting out at PL2 in Grade 3 (N = 28), four paths - *stability* (29%), *increase/decrease* (25%), *early increase* (18%), and *late increase* (11%) - represented approximately 80% of these students’ data. For children starting out at the PL1 in Grade 3 (N = 13), three paths - *stability* (31%), *increase/decrease* (31%), and *late increase* (23%) - represented approximately 80% of these students’ data, with the path of *two category change* required to account for the remaining data.
In the next analysis, the proportion of students within each of the Grade 10 achievement levels that had followed each of these developmental paths (Table 2.7) was examined. For students classified as poor readers in Grade 10 (N = 14) the vast portion of the data (86%) was equally accounted for by three different paths (stability, increase/decrease, and two level change). For Grade 10 below-average readers (N = 25), two paths (stability and increase/decrease) equally accounted for 64% of the data. For Grade 10 average readers (N = 33), stability accounted for 52% of the data. One third of students classified as above-average readers (N = 6) followed a path of early increase.

Less than half (38.5%) of the students in the total sample exhibited stability for all three periods. In contrast, 82% of students exhibited stability of Grade 10 reading achievement categorization with either Grade 3 or 6 categorization (stability of individual Grade 10 achievement groupings was lowest for PL4 at 67% and highest for PL3 at 88%).
Table 2.7

*Prevalence (Percentages) of Development Paths from Grades 3 and 6 Based on Grade 10 Reading Achievement*

| Developmental Path   | Poor  
|                      | Below-Average  
|                      | Average  
|                      | Above-Average  
<table>
<thead>
<tr>
<th></th>
<th>Total Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stability</td>
<td>28.6</td>
</tr>
<tr>
<td></td>
<td>32.0</td>
</tr>
<tr>
<td></td>
<td>51.5</td>
</tr>
<tr>
<td></td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td>38.5</td>
</tr>
<tr>
<td>Early Increase</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>15.2</td>
</tr>
<tr>
<td></td>
<td>33.3</td>
</tr>
<tr>
<td></td>
<td>9.0</td>
</tr>
<tr>
<td>Late Increase</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td>9.0</td>
</tr>
<tr>
<td>Early Decrease</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>9.0</td>
</tr>
<tr>
<td>Late Decrease</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>2.6</td>
</tr>
<tr>
<td>Increase/Decrease</td>
<td>28.6</td>
</tr>
<tr>
<td></td>
<td>32.0</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>15.4</td>
</tr>
<tr>
<td>Decrease/Increase</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td>9.0</td>
</tr>
<tr>
<td>Two Category</td>
<td>28.6</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td>7.7</td>
</tr>
</tbody>
</table>
**Regression toward the mean.** Whenever change over time is analyzed, it is important to address issues that arise from regression effects. Regression toward the mean is the statistical trend for scores at one time to move closer toward the mean at a second time, and is often described in terms of errors of measurement (Campbell & Stanley, 1963). It is observed whenever the correlation between two variables is less than perfect for whatever reasons (Furby, 1973), and the lower the correlation between the scores at the two times, the greater the regression toward the mean. For the present purpose, two tests were used to study the regression effects in the data. Following Phillips et al. (2002), a time-reversed control analysis, recommended by Campbell and Stanley (1963), and a comparison between predicted scores based on regression coefficients derived from the entire sample and the actual scores observed in the subgroups, suggested by Lord (1958), were used.

The time-reversed analysis (cf. Phillips et al., 2002) involved calculating the later mean scores of students who had been grouped into the four achievement levels at an earlier time, and comparing the results with the earlier mean scores of students who had been grouped into the four achievement levels in the later period. If regression toward the mean were operating, children scoring above the mean at the earlier time would obtain lower scores at the later testing time (a downward sloping time-forward line) and children scoring below the mean in the earlier period would obtain higher scores at the later testing time (an upward sloping time-forward line); the opposite trend would be expected in the time-backward analysis. Time-reversed analyses were conducted for the Grade 3 to 6, Grade 6 to 10, and Grade 3 to 10 periods (see Figure 2.1). Considering the forward time analyses (solid lines), in all three periods the higher-level groups (PL3 and PL4) exhibited a decline in achievement level whereas the lower level groups (PL2 and PL1) exhibited an increase, as would be expected due to regression toward the mean. Figure 2.1, however, also shows the time-reversed control
analysis (broken lines), and in general reveals the opposite picture. Subjects scoring higher at the later time exhibited a relative increase from the earlier time, and those scoring lower at the later times exhibited a performance decrease from the earlier time, again as would be expected in regression toward the mean. In general, the magnitude of the forward changes is similar to those of the backward changes, with several exceptions. The largest of these discrepancies is for PL4 in Grade 6 to 10, for which there is less regression toward the mean than is shown in the backward analysis. The complementary pattern is shown for the PL1 students in Grade 6 to 10. Thus, the time-reversed analyses suggest that the majority of the observed changes were due to regression effects, and most of the exceptions are due to stability.

Figure 2.1

*Comparison of Time-forward (solid lines) with Time-backwards Analysis (broken lines) for each Achievement Group for Grade 3 and 6, Grade 6 and 10, Grade 3 and 10*

In the second analysis, regression analysis was used to examine the difference between predicted and actual trends (cf. Campbell & Stanley, 1963). Predicted values of later grade scores from earlier grade scores were calculated by regression. Actual values were derived by calculating the mean later achievement for students classified in each earlier achievement group. Given the imperfect correlation across years (see Table 2.2), students
with achievement above the mean in the earlier year should regress downward at the later time, and those with achievement below the mean in the earlier year should regress upward at the later time. If the actual changes are similar to the predicted changes, this would indicate that they may be due to regression effects. Figure 2.2 presents the results based on separate regression analyses for each of the three periods examined. With three exceptions, the difference between the actual and predicted scores was small. The actual, predicted, and difference scores (the last expressed as a fraction of one achievement level) are shown in Table 2.8. The first exception was in the prediction of Grade 6 scores from Grade 3 scores, the actual value for PL2 students was .23 of an achievement level higher than predicted by regression, suggesting real improvement. The second exception was in the prediction of Grade 10 scores from Grade 6 scores, the actual Grade 10 value for PL4 students ($N = 5$) was .40 of an achievement level higher than predicted. This indicates more stability than expected from regression. The third exception was in the prediction of Grade 10 scores from Grade 6 scores: the PL1 students showed less gain than would be predicted from regression (difference of .18), again indicating more stability than predicted from regression. Thus, only one of the exceptions indicates more growth than would be expected by regression. Furthermore, it is worth noting that the largest difference in the Grade 3 to 10 analysis was .13 of a level, indicating the PL3 students showed more stability than expected.

Both techniques indicate that the majority of the apparent change across grades could be due to regression toward the mean. In the exceptions to this pattern, most demonstrated more stability than expected. There was very little evidence of improvements by lower-performing students.
Figure 2.2

Progress of Children based on Actual (solid lines) and Predicted Scores (dotted lines) of Achievement Groupings for each Time Period

![Diagram showing progress of children based on actual and predicted scores](image)

Table 2.8

Actual (A), Predicted (P) and Difference (D) Scores for each Time Period

<table>
<thead>
<tr>
<th>Level</th>
<th>Grade 6 actual and predicted scores based on Grade 3 to 6 regression</th>
<th>Grade 10 actual and predicted scores based on Grade 6 to 10 regression</th>
<th>Grade 10 actual and predicted scores based on Grade 3 to 10 regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>P</td>
<td>D</td>
</tr>
<tr>
<td>PL4</td>
<td>3.20</td>
<td>3.26</td>
<td>-.06</td>
</tr>
<tr>
<td>PL3</td>
<td>2.72</td>
<td>2.70</td>
<td>.02</td>
</tr>
<tr>
<td>PL2</td>
<td>2.36</td>
<td>2.13</td>
<td>.23</td>
</tr>
<tr>
<td>PL1</td>
<td>1.54</td>
<td>1.57</td>
<td>-.03</td>
</tr>
</tbody>
</table>
Discussion

This study investigated the extent to which the stability, cumulative, and compensatory models describe the development of reading comprehension. Measures of students’ reading achievement for three time points (Grades 3, 6, and 10), and four measures of reading comprehension were used (one measure for each of the Grade 3 and 6 testing points, and two for Grade 10). Three main questions were addressed. The first question concerned the relationships among the four comprehension measures, the second the conditional probabilities of later reading achievement given earlier level of reading achievement, and the third the developmental paths of reading achievement from Grades 3 to 6 and 10, and their prevalence.

The magnitude of the correlations among the reading comprehension measures indicates a moderate but substantial relationship between measures and across times (range .54 to .65). These correlations are consistent with those reported in other studies based on broad samples of children. For example, Cunningham and Stanovich (1997) reported a Grade 1 to 11 correlation of .58 using two different standardized tests of reading comprehension, and Phillips et al. (2002) a Grade 1 to 6 correlation of .59 using the same standardized reading comprehension test. With regard to the two Grade 10 measures, their correlation (.60) was in the lower range reported in other studies that have used more than one concurrently administered reading comprehension measure. Cutting and Scarborough (2006), for example, reported correlations of .64 to .79 for three commercially available tests with age controlled. However, my results are consistent with Rupp and Lesaux’s (2006) correlation in Grade 4 of .64 between a commercially available measure (Stanford Diagnostic Reading Test) and a government standards-based assessment of reading comprehension. Generally, the trend for lower correlations is to be expected when a limited number of achievement groupings (as were used in this study and by Rupp and Lesaux)
rather than raw scores (as by Cutting and Scarborough) are used. The correlation (.55) between
the Grade 3 and 6 government-mandated assessments is comparable to that between two
government-mandated assessments reported by Carreker et al. (2007) in a Grade 3 to 5 sample
(.60). Overall, the present results support the findings of many studies indicating substantial
stability in reading comprehension achievement, yet still leaving considerable room for
developmental changes.

The conditional probability and developmental paths analyses allowed us to explore in
more detail the extent and nature of these developmental changes. Specifically, my review of the
literature suggested that students may follow different paths in their reading development –
stability, cumulative, and compensatory. Previous studies (e.g., Leppänen et al., 2004; Parrila et
al., 2005; Phillips et al., 2002) have found that some children who start with lower scores may
follow a compensatory path in their reading development. Some initial support for the
compensatory path of some low-starting students was obtained from the results of my
conditional probabilities – that is, a substantial portion of low-starting students (defined in this
study as achieving PL1 and PL2) improved their achievement by a later time. For example, 20%
to 40% of those at PL1 in an earlier period moved to PL2 at a later time (see Tables 2.3, 2.4, and
2.5). However, this initial support was tempered by my tests investigating statistical effects
associated with regression toward the mean. These results indicated that real compensation
applied to the Grade 3 to 6 period only, and was limited to a small percentage of PL2 students.
Importantly, these PL2 starting students would have been classified as average readers rather
than poor readers had the commonly-used cut-off of one standard deviation below the mean been
used to establish low achievement groupings. Evidence that the lowest starting students (PL1
students representing the bottom 17% of my sample) demonstrated real compensation was not found. Differences in sample, measures, context, and curriculum between studies may explain some of the dissimilarity in findings.

Our review of the literature suggested contradictory findings regarding the cumulative model. With regard to relative decreases for the achievement of low-starting students, a low probability for the performance by PL2 students to decrease was found. Students starting at PL1 could go no lower: data such as these cannot be sensitive to declines within this level. Previous research has shown a relative decline in reading achievement for some students with initially adequate achievement (e.g., Leach et al., 2003). This has been most commonly investigated with students as they progress from the primary to middle school years. Results for the conditional probabilities for Grade 3 to 6 appeared to provide support for this previously observed phenomenon. Further, my results for the conditional probabilities for Grade 6 to 10 suggested that there is a group of students achieving adequate performance in Grade 6 who experience a relative decline in Grade 10. However, my analyses for the developmental paths using three time points indicated that the relative achievement in Grade 10 was for the most part stable with their initial Grade 3 achievement. To emphasize this point, more than half of the students who achieved PL3 in Grade 3 and who experienced a relative decline in performance in the Grade 6 period returned to their Grade 3 achievement category in Grade 10. Further, my results indicate that the relative decreases observed in some students who began at PL3 could be largely explained by regression toward the mean. Unlike Phillips et al. (2002) who reported a relative decline in performance in their initially above-average group that could not be entirely explained by regression toward the mean, I did not find such evidence. Indeed, for children classified as
PLA in Grade 6, their actual achievement in Grade 10 was slightly higher than would have been predicted. Again, differences in findings across studies may be accounted for by variables mentioned above.

There are many potential explanations for the stability observed in students’ relative achievement. A large body of research suggests that a strong candidate to explain poor reading comprehension performance is weakness in word reading skill and in the skills that support word reading (phonology, orthography, morphology, and vocabulary; e.g., Kirby, Desrochers, Roth, & Lai, 2008). Further, because students who do not read well generally do not read very much, they have fewer opportunities for reading practice and for learning from what they have read (Anderson, Wilson, & Fielding, 1988). Lower reading skill and less practice lead to a decrease in motivation to read, which further acts to counter remedial efforts. At the upper end of the distribution, these same factors work to maintain students’ initial advantages in reading comprehension: increased skill leads to more reading practice, to greater vocabulary, to more content knowledge, and to greater motivation to read. These influences are those described in the cumulative model (Leppänen et. al., 2004; Stanovich, 1986). Although not much evidence was found in this study for the cumulative model, that may be largely due to the nature of the data. The four-category scoring system did not allow for the students in the top group to improve or those in the bottom group to decline. The factors implicated in the cumulative model can also be seen to support stability.

Limitations

There are certain limitations to the present study that need to be acknowledged. First concerns issues regarding the nature and size of my sample. English was the primary language
for all of the children included in my sample, and so my data should not be extrapolated to populations of children whose primary language is not English. My sample included students attending one Canadian high school thereby limiting the generalizability of my findings to other populations in other countries. My sample was limited to only 78 students with the result that some of the subgroup analyses were underpowered. Second, the categorical rather than continuous nature of the data, and the small sample size precluded the use of newer multilevel analyses (e.g., latent growth curve analysis). Third, I utilized a retrospective design; therefore, additional measures assessing basic reading (e.g., word reading, decoding), vocabulary, language were not available to explore why a few children seemed to follow different pathways besides stability. Fifth, only one measure of reading achievement was included for the Grade 3 and 6 assessments. The inclusion of a composite score of reading comprehension for Grade 10 achievement may have reduced the magnitude of the statistical artifact of regression towards the mean.

Implications and Conclusions

The finding of considerable stability in students’ reading comprehension development has several implications, especially for less able readers. First, the majority of those demonstrating poor reading ability in Grade 10 were identified as such as early as Grade 3; it is possible that other testing could have led to even earlier identification. Second, if this is so, then early intervention is warranted. Previous intervention studies (e.g., Fuchs & Fuchs, 2006; Pinnell, 1989; Wasik & Slavin, 1993) indicated that a substantial portion of reading-impaired children can acquire at least grade-level reading skills if they receive early and intensive
remediation. Remediation that is delayed until children are older may be more difficult (Roberts, Torgesen, Boardman, & Scammacca, 2008; Shaywitz, Morris, & Shaywitz, 2008).

Effective reading comprehension, however, depends upon a host of language-based and cognitive skills that go beyond those involved in accurate and fluent word reading. Reading comprehension difficulties may stem from inadequate vocabulary or conceptual knowledge, weak reasoning or inferential skills, lack of motivation, or an inability to apply active comprehension strategies (Cunningham & Stanovich, 1997; Guthrie et al., 2007; Kirby & Savage, 2008; National Reading Panel, 2000). Non-instructional variables such as family stability and mobility, parental involvement and expectations, parent education level, and student capacity for learning have all been shown to influence test performance (Simner, 2000). The influence of such factors has been found to make it difficult to improve weaknesses in early reading achievement (Moss et al., 2008).

Although the overall finding of stability may seem less exciting than one showing dramatic changes, it should not obscure the fact that all students are improving in their reading comprehension. Overall stability on average does not exclude the possibility of dramatic individual gains, even though few of these were observed in this sample; it should also be remembered that the four-point scoring scheme used in these measures is less than ideal for observing fine-grained details of change, or to applying covariance modeling techniques for removing statistical effects resulting from measurement error (Ferrer & Zhang, 2009). Furthermore, the observed stability should not be taken to imply that change is impossible: the results simply reflect the way matters stand at the moment. More intense, focused, and earlier intervention may produce greater improvement. I conclude that reading comprehension
difficulties can be identified early in students’ school careers, and that this may be the best time to intervene.
References


Simner, M. L. (2000). A joint position statement by the Canadian Psychological Association and the Canadian Association of School Psychologists on the Canadian press coverage of the


CHAPTER 3: EXAMINING THE SPECIFICITY OF COMPREHENSION DIFFICULTIES IN UNEXPECTED POOR COMPREHENDERS

Abstract

This study examines the specificity of the comprehension weaknesses of 15 year old readers whose comprehension skills are below those expected from their skill in word reading and nonverbal ability (unexpected poor comprehenders). Regression analyses identified unexpected poor comprehenders, and two contrast groups (expected average and unexpected good comprehenders). Characteristics of unexpected poor comprehenders are examined after controlling for word-reading accuracy, phonological decoding, reading rate, nonverbal ability, and vocabulary. Findings indicate a critical disadvantage of unexpected poor comprehenders lies in their weakness in vocabulary and that comprehension difficulties related to the identification of details and main ideas in summary writing remain when vocabulary is controlled. Implications for interpreting previous and informing future research are discussed.

Introduction

Although most reading comprehension difficulties are associated with difficulties in decoding, some readers demonstrate poor reading comprehension even though they have adequate decoding skills (Cain & Oakhill, 2007). A vast amount of research has been devoted to examining comprehension difficulties that arise because of poor decoding skill (for a review see Vellutino, Fletcher, Snowling, & Scanlon, 2004). Less research has been devoted to investigating the sources of comprehension difficulties of children who possess age appropriate skill in word reading, but who still have trouble comprehending what they have read. By some estimates, 10% of 7 to 10 year-old children present this profile (Leach, Scarborough & Rescorla, 2003; Nation &
Snowling, 1997; Yuill & Oakhill, 1991). Studies with older populations indicate a higher frequency (15% to 20%; Catts, Adlof, Hogan, & Ellis Weismer, S., 2006; Landi & Perfetti, 2006). These students have been referred to as poor comprehenders (Stothard & Hulme, 1992), less-skilled comprehenders (Ehrlich & Remond, 1997; Nation, Marshall, & Altman, 2003), or as having a specific comprehension problem (Oakhill, Hartt & Samols, 2005). The age-appropriate word reading ability of this group of children, coupled with their relatively weaker reading comprehension, indicates that their performance can be considered to be unexpectedly poor. Throughout this paper I will describe them as unexpected poor comprehenders, to distinguish them from other readers whose poor comprehension is consistent with their poor word reading skills.

**Introduction**

In this paper I focus on the specificity of the comprehension weaknesses of these unexpected poor comprehenders and examine both lower-level word-related skills and knowledge and higher-level cognitive processes that have been related to comprehension in order to understand more about reading failure in this group. I first describe the overall findings of studies carried out with unexpected poor comprehenders, and then discuss theoretical considerations for investigating reading comprehension difficulties. In light of these theoretical considerations I then review in more detail the findings and limitations of studies carried out with this group of poor comprehenders. I then present the results of a study designed to address these limitations and increase our understanding of unexpected poor comprehenders.

**Unexpected Poor Comprehenders**

Previous research investigating the source of comprehension difficulties of unexpected poor comprehenders has reported weaknesses, relative to more skilled comprehenders, on a
range of semantic tasks. In studies that have controlled for age, pseudoword reading, and nonverbal ability, unexpected poor comprehenders, relative to more skilled comprehenders, have been found to (a) generate fewer semantic category members in a verbal fluency task (Nation & Snowling, 1998a), (b) have selective difficulty for abstract or low imageability words relative to concrete words in semantic judgment and recall tasks (Nation & Snowling, 1999), (c) be less able in naming pictures with less frequent labels (Nation, Marshall, & Snowling, 2001), and (d) perform more poorly on a variety of tasks assessing their vocabulary knowledge (Nation & Snowling, 1998b; Ricketts, Nation, & Bishop, 2007). These findings suggest that the source of comprehension difficulties of unexpected poor comprehenders rests with semantic processing at the word level (Nation & Snowling, 1998a, 1999). Slowness or inability to access word meanings affects processing at higher levels, be it at the sentence or text level (Perfetti, 1985).

In other studies that have controlled instead for age, word reading accuracy, and vocabulary, weaker performance by unexpected poor comprehenders, relative to more skilled comprehenders, has been observed on tasks that require readers to make (a) inferences that elaborate on text information using prior knowledge (Cain, Oakhill, Barnes, & Bryant, 2001), (b) inferences that require the integration of information provided in the text with general knowledge (Cain & Oakhill, 1999), and (c) use of the context of a story to infer the meanings of single novel words, with these difficulties particularly pronounced when the novel word and the useful contextual clues were separated by filler text (Cain, Oakhill, & Elbro, 2003; Cain, Oakhill, & Lemmon, 2005). These findings suggest that this group has difficulties with the higher-order comprehension skills of inferencing, text integration, and integrating text information with their background knowledge (Oakhill, 1984; Cain et al., 2001; Cain & Oakhill, 1999). Based on
correlational evidence, both lines of research indicate unexpected poor comprehenders, relative to more skilled comprehenders, have semantic weaknesses. However, given that these studies have differed in the control procedures used, the theoretical basis for this group’s comprehension difficulties has not been clearly defined.

**Theoretical Considerations**

Current theories stress the multi-level nature of comprehension (e.g., Graesser, 2007; Kintsch & Kintsch, 2005; Rapp, van den Broek, McMaster, Panayiota, & Espin, 2007). Word-level skills are required (Perfetti, 2007), but effective word reading is by no means a guarantee of comprehension (Kintsch, 1998). The nature of lower-level word reading skill is outlined in the lexical quality hypothesis (Perfetti & Hart, 2001; Perfetti, 2007). It proposes that skill in reading comprehension requires efficient word identification skills that are supported by readers’ knowledge of the form and meaning of words, including the precision of the representation of the constituents of “orthography, phonology, morphology and meaning” (Verhoeven & Perfetti, 2008, p. 294). Accordingly, comprehension may fail either because readers lack the necessary word representations, or because the quality, including diversity, efficiency, and integration, of lexical representations is low.

The focus of Kintsch’s construction-integration model (1998) is on the higher-level comprehension processes that take place after decoding processes and the semantic processes of attaching relevant word meanings have been completed successfully. The first level of comprehension is at the surface level, in which propositions are formed. The next level of comprehension occurs at the textbase level, in which propositions are organized in terms of importance at the micro- and macro-propositional levels. The highest level of comprehension is
the situation model, a structure which combines text information with prior knowledge to support inferences and abstraction.

Thus comprehension can be seen to occur at several levels. Word-level processes concern the quality of word representations (Perfetti & Hart, 2001; Perfetti, 2007), whereas text-level processes concern the construction of representations of text meaning at several levels (Kintsch, 1998). These theoretical distinctions are useful in reviewing findings from existing studies of unexpected poor comprehenders.

**Previous Research with Unexpected Poor Comprehenders**

**Word Knowledge**

The literature shows no evidence that weak phonological skills are a source of impairment for unexpected poor comprehenders. Regardless of whether nonword reading and nonverbal ability, or word reading and vocabulary have been controlled, unexpected poor comprehenders have not been found to differ from skilled comprehenders on measures assessing phonological processing skill (i.e., timed or untimed pseudoword reading, or phoneme deletion tasks) (e.g., Cain, Oakhill, & Bryant, 2000; Catts et al., 2006; Nation & Snowling, 1998a; Nation, Clark, Marshall, & Durand, 2004; Ricketts et al., 2007). This is consistent with the prevailing theory that phonological awareness and phonological decoding contribute to word reading (e.g., Share, 1995), so that once word reading skills are controlled, no further differences should be apparent in phonological skills.

Fewer studies have been carried out assessing unexpected poor comprehenders’ orthographic knowledge. In those that have, regardless of whether nonword reading and nonverbal ability (Ricketts et al., 2007), or word reading, reading rate, and nonverbal ability (Tong, Deacon, Kirby, Cain, & Parrila, 2011) have been controlled, they have not been found to
differ from more skilled comprehenders in orthographic processing. Again this is consistent with the notion that orthographic skills make their primary contribution to word reading (Coltheart, Rastle, Perry, Langdon, & Ziegler, 200), and so should not show differences once word reading is controlled.

The literature is less clear regarding the morphological skills and knowledge of poor comprehenders. Nation and colleagues (2004) reported that 8 year old unexpected poor comprehenders, matched with more skilled comprehenders for decoding and nonverbal ability, had difficulties on measures of morphosyntax (elicitation of regular and irregular past tense verbs and recall of sentences of increasing complexity). However, as Nation et al. pointed out, weaknesses in their vocabulary and with higher level oral language comprehension were also found, thereby placing their observed weaknesses in morphology into a broader set of language weaknesses. Results from a longitudinal study (grades 3 to 5) reported by Tong et al. (2011) in which context-free word reading accuracy, reading rate and nonverbal ability were controlled, found unexpected poor comprehenders to have selective deficits on tasks assessing morphological awareness. In this study, the performance of unexpected poor comprehenders was contrasted with both average and above-average comprehenders on derivational and inflectional tasks. Poor comprehenders performed significantly lower on the derivation task in grade 3 than the above-average group, but not the average group. In grade 5 poor comprehenders performed significantly lower than both contrast groups on derivations. These differences in morphological skills are consistent with a model in which morphological skills make additional contributions to reading comprehension beyond those they make to word reading (Kirby, Deacon, Bowers, Izenberg, Wade-Woolley, & Parrila, 2011).
With regard to the role that vocabulary skill plays in accounting for the reading comprehension difficulties of unexpected poor comprehenders, two different lines of research have been pursued. One has relied on pseudoword reading to control for reading accuracy, in addition to nonverbal abilities, and has found that unexpected poor comprehenders, compared to more skilled comprehenders, exhibit weaknesses in vocabulary (e.g., Cragg & Nation, 2006; Nation & Snowling, 1998b; Ricketts et al., 2007). This finding suggests that unexpected poor comprehenders’ weaknesses in vocabulary skill contribute to their reading comprehension difficulties. But interpretation of these results is not entirely clear as in some of these studies the word reading skills of the unexpected poor comprehenders, in addition to their vocabulary skills, have also been weaker than those of more skilled comprehenders (e.g., Cragg & Nation, 2006; Nation & Snowling, 1998b). Allowing both word reading skills and vocabulary to vary makes it difficult to determine the level at which vocabulary skill contributes to reading comprehension. Indeed, differences in vocabulary skill are consistent with a model in which vocabulary skill makes additional contributions to reading comprehension skill beyond those made to word reading skill (e.g., Goff, Pratt, & Ong, 2005).

In the second line of research, vocabulary has been controlled in addition to word reading accuracy. Findings from these studies indicate that unexpected poor comprehenders exhibit difficulties with the higher-order comprehension skills of inferencing, text integration, and integrating text information with their background knowledge (Oakhill, 1984; Cain et al., 2001; Cain & Oakhill, 1999). These findings suggest that vocabulary skill does not account for the comprehension difficulties of unexpected poor comprehenders, and are consistent with the acknowledgement that some, but not all reading comprehension difficulties can be attributed to
weaknesses in word knowledge (Perfetti, 2007). However, what is presently unclear from this second line of research concerns the specific contribution that vocabulary skill makes to reading comprehension skill. To this end, an investigation that assesses the effect of controlling vocabulary is necessary.

**Textbase and Situation Model Comprehension**

Beyond word-level knowledge, Kintsch’s construction-integration model (1998) distinguishes between two higher levels of comprehension. At the next level, textbase comprehension is necessary with a distinction made between micropropositions (details) and macropropositions (explicitly stated main ideas). Further processing is required for comprehension at the level of the situation model. This level can involve inferring main ideas and thematic elements of a text. The distinctions made in the construction-integration model between levels of comprehension raise the question whether unexpected poor comprehenders have weaknesses at some, or all of these levels.

The findings regarding the micropropositional level, represented by literal comprehension performance, have not been conclusive. Some studies have reported that unexpected poor comprehenders perform more poorly, compared to more skilled comprehenders, on questions assessing literal comprehension (Cain et al., 2001; Oakhill, 1984), but other studies have not (Cain & Oakhill, 1999, Bowyer-Crane & Snowling, 2005; Oakhill, 1982). Although there is considerable evidence that unexpected poor comprehenders have deficits in inference making (see Cain & Oakhill, 2007, for a review), it is not clear to what extent those effects have been at the micropropositional as opposed to higher levels. Other evidence at the macropropositional level has not concerned reading comprehension directly. For example, several studies have
assessed 7 and 8 year old poor comprehenders’ skill in narrative expression (Cragg & Nation, 2006; Cain, 2003). Poor comprehenders have been found to produce stories that have fewer main ideas, fewer inferred causal relations (Cragg & Nation), and less integrated event structures (Cain, 2003) than more skilled comprehenders. However, in both of these studies reading comprehension was not explicitly assessed and children only responded to picture, title, or topic prompts. The results provide some evidence of weaknesses at the macropropositional level, but further evidence is required regarding memory representations obtained from reading.

**Other Factors Contributing to Individual Differences in Reading Comprehension**

It is also necessary to consider other skills and processes that contribute to effective reading comprehension. They include, but are not limited to, individual differences in working memory (Daneman & Merikle, 1996), the use of active strategies (Cain, 1999; Pressley & Afflerbach, Fuchs & Fuchs, 2005; Samuelstuen & Bråten, 2005; RAND Reading Study Group, 2002), and print exposure (e.g., Bolger, Balass, Landen, & Perfetti, 2008; Cunningham & Stanovich, 1997; Landi, 2010; Perfetti & Hart, 2001, 2002; Stanovich, 1986).

A meta-analysis of 18 studies of individuals with specific reading comprehension difficulties reported a .75 effect size difference between unexpected poor and good comprehenders (matched on measures of decoding and cognitive abilities) for complex verbal memory span tasks (Carretti, Borella, Cornoldi & De Beni, 2009). However, considerable heterogeneity in effect size across studies was also reported by the authors. They suggested that this may have been due to the different procedures adopted for measuring working memory. The heterogeneity may also have been due to the wide variation across studies in the control procedures that were used. Some of the studies relied on nonword reading, others on contextual
word reading, and still others on context-free word reading accuracy to control for word reading skill. The studies also differed in the additional control variables that were used; some also controlled verbal ability, others nonverbal ability. None of the studies included in the meta-analysis included controls for both verbal and nonverbal ability.

Findings regarding the print exposure of unexpected poor comprehenders have been equivocal. Some studies have reported them to have less reading experience than more skilled comprehenders (an unpublished study by Cain, cited in Oakhill & Yuill, 1996; Torppa, Tolvanen, Poikkcus, Ekiund, Lerkkanen, Leskinen, & Lyytinen, 2007), and others not (Ricketts et al., 2007; Cain et al., 2000). The findings of these studies have been based on children 7 – 10 years of age. As children grow older, it is possible that the weaker comprehension skills of unexpected poor comprehenders result in their deriving less pleasure from reading and therefore spending less time reading. The converse may be true of more skilled comprehenders. This, however, remains to be investigated.

Findings from studies investigating individual differences in strategy use between unexpected poor comprehenders and more-skilled comprehenders have not been conclusive. Two studies (Cain, 1999; Catoldo & Oakhill, 2000) have reported that unexpected poor comprehenders rely on surface strategies (such as focusing on the reading or remembering of individual words). In contrast, Yuill and Oakhill (1991) found that both good and poor comprehenders characterized a superficial reader as “one who can read lots of long words”, and a deep reader as “one who knows what things mean.” Differences in samples and the contexts in which specific strategies were assessed (use vs. beliefs) may explain such contradictory findings. Differences in strategy use between more- and less-skilled comprehenders may also be due to
differences in nonverbal ability, a variable which has not always been controlled (Cain, 1999, Cataldo & Oakhill, 2000), and which, in turn, may be related to other variables (e.g., working memory capacity; Salthouse & Pink, 2008).

**Present Study**

I suggest that a clearer understanding of the comprehension difficulties of unexpected poor comprehenders will benefit from studies that address the following six limitations or gaps I have identified in previous research. The first concerns the measure of word reading employed in matching groups. Many of the previous studies investigating unexpected poor comprehenders have relied on a reading measure (Neale Analysis of Reading Ability: NARA) in which word reading and reading comprehension abilities are not assessed independently: children’s word reading accuracy is assessed while reading connected text, and mispronounced or omitted words are supplied by the tester (Cain, 1999; Cain & Oakhill, 1999; Cain et al., 2004, 2005; Nation, Adams, Bowyer-Crane, & Snowling, 1999; Nation & Snowling, 1998a, 1998b; Oakhill, 1982; Oakhill, Yuill, & Parkin, 1986; Stothard & Hulme, 1992, 1995; Yuill, Oakhill, & Parkin, 1989). Reading words in context has been shown to have a bootstrapping effect on word reading ability (Perfetti & Hogaboam, 1975; Nation & Snowling, 1998b), and thereby may overestimate the word reading abilities of those with good comprehension, and, conversely, underestimate those of students with poor comprehension. On the other hand, supplying words that have not been read correctly may allow readers to comprehend texts that they otherwise could not have comprehended. In attempting to disentangle the effects of word reading and reading comprehension, I suggest that it is important to use a context-free measure of word reading (i.e., isolated word reading accuracy); at the very least, it is important to ensure that similar results are found with such a measure. Furthermore, accurate word reading does not rule out the possibility
that the word reading skills of unexpected poor comprehenders are slower or less efficient than those of more skilled comprehenders. Controlling for reading fluency, in addition to reading accuracy, has not been routinely included in studies of unexpected poor comprehenders. In the present study I control word reading ability with measures of isolated real word reading accuracy and word reading rate. Because word reading skill is closely associated with skill in phonological decoding (Vellutino et al., 2004), pseudoword reading is also included as a control measure of reading skill. As outlined in the lexical quality hypothesis, effective reading skill is supported by orthography, phonology, morphology and meaning (Verhoeven & Perfetti, 2008). In the present study, measures assessing each of these word related skills are included.

My second concern is which cognitive abilities to control. In addition to word reading ability, the majority of previous studies have matched unexpected poor comprehenders with more skilled comprehenders on either nonverbal ability (Nation & Snowling, 1999), or written vocabulary knowledge (Cain, 2006; Oakhill, 1982). Controlling for vocabulary is important because it allows us to distinguish between semantic weaknesses at the level of word knowledge and those at the discourse-level of comprehension (i.e., textbase and situation model). Controlling nonverbal ability prevents confounding reading comprehension differences with differences in general cognitive ability and may eliminate other factors that are strongly related to nonverbal ability (such as working memory; Salthouse & Pink, 2008). I therefore argue that the best evidence for understanding the deficits of poor comprehenders comes from controlling, in addition to word reading skill, both vocabulary and nonverbal abilities. Controlling each in turn will also allow us to determine what effect each has upon which participants are identified as
unexpected poor comprehenders. Other relevant individual difference factors (i.e., working memory capacity, strategy use, and print exposure) will also be included.

The third gap that I see concerns the theoretical distinction between difficulties occurring at different levels of comprehension (i.e., the word level or vocabulary, micropropositions, macropropositions, and the situation model). As discussed earlier, previous research has either yielded contrary results (e.g., for literal comprehension, Cain et al., 2001, and Oakhill, 1984, versus Cain & Oakhill, 1999), or has not made the distinction between levels. In this study, I include a vocabulary measure and reading comprehension measures assessing understanding at various levels.

The fourth limitation concerns the way in which unexpected poor comprehenders have been identified. Use of cut-off scores has several unfortunate consequences, including likely mis-categorization of participants near the cut-off boundaries and omission of participants who have a relative discrepancy but do not meet one of the cut-off criteria (e.g., ones with very high word reading but average comprehension). I employ the regression method to identify participants that was used by Tong et al. (2011), and which is described more fully in the Results section. The regression technique allows us to include a range of reading skills of age-matched participants, and to explore the consequences of controlling different variables.

My fifth concern is about the group(s) to whom unexpected poor comprehenders have been compared. Most previous research has compared poor comprehenders to one contrast group (e.g., Cain & Oakhill, 2006; Oakhill, 1982; Nation & Snowling, 2000; Yuill et al., 1989) rather than to several groups which differ in reading comprehension skill (e.g., Cain & Oakhill, 1999; Nation, Marshall, & Snowling, 2001; Catts et al., 2006; Tong et al., 2011). It is possible that
some of the differences observed between poor comprehenders and the single, broadly-based comparison group are due to the high performance of participants with very high comprehension skills, much higher than would be expected from their word reading skills. In searching for deficits that account for their unexpected poor comprehension, it is more reasonable to compare them to average comprehenders. This is what is done in this study, though I include a group of unexpected high comprehenders to observe what effect they may have had in previous studies.

The last issue concerns the age group studied, with the majority of studies carried out with 7 to 10 year-old participants. Developmental growth in word reading and comprehension skill, however, continues as children proceed in their schooling (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996). At the high school level, students face increasingly complex narrative and expository texts, with progressively greater demands on their comprehension skills. Students with weak comprehension skill will have difficulty meeting the demands of these more complex texts which are required for them to successfully complete their coursework requirements (Kamil, 2003; Snow & Biancarosa, 2003). To extend the research towards an older group of students, my study is carried out with 15 year-old students.

**Research Questions**

The purpose of this study is to investigate the sources of reading difficulties of unexpected poor comprehenders in high school. My research questions are meant to address the gaps and limitations of previous research discussed above. The study addresses whether (a) different control measures lead to the identification of different individuals; (b) higher-level semantic weaknesses in textbase and situation model comprehension remain when word reading, phonological decoding, reading rate, vocabulary and nonverbal ability are controlled; (c) any
observed weaknesses can be attributed to weaknesses in performance on word-related tasks (i.e., phonological awareness, orthographic knowledge, and morphological awareness), or individual differences variables (i.e., working memory, approaches to learning, and exposure to print); and, (d) any observed differences depend upon the nature of the contrast used (average versus above-average comprehenders). Addressing these questions will assist in interpreting the sources of weaknesses reported in previous studies of unexpected poor comprehenders.

Method

Participants

The participants were 137 Grade 10 students enrolled in a mid-sized rural Ontario Catholic high school (mean age 15.0 years, SD 6.3 months; 78 girls and 59 boys). Students attending the school represented a range of socio-economic backgrounds. A comparison of demographic characteristics of the district served by the school to that of the province (Statistics Canada, 2001) indicates greater employment in manufacturing and construction (34% versus 22%), a greater portion with less than a high school diploma (28% versus 20%), and a smaller proportion with university education (11% versus 24%). Excluding those people with an aboriginal identity (9% versus 2%), there were very few visible minorities (1% versus 19% provincially). English was the first language learned by a large segment of the population (88% versus 71% provincially).

Participants were enrolled in one of two streams (63.5% academic and 36.5% applied) of Grade 10 English classes. Whereas the academic stream is intended to prepare students for university, the applied stream is intended to prepare students for post-secondary education at a vocational college or for direct entry into work after high school. In the present sample, 15% of the students (N = 20) had been identified with a special learning need; 15 of the students were formally identified by the school board (12 learning disabled, 1 gifted, 1 behaviour, 1 mild
intellectual disability), and 5 informally identified with only an Individual Education Plan (1 mild hearing loss and 4 unspecified). Student and parental permission for participation was obtained (see appendix A).

**Measures**

**Demographic questionnaire.** Participants completed questions regarding their gender, level of academic study, first language learned, and age.

**Control measures.** Three measures of reading, one of nonverbal ability, and one of vocabulary were administered.

**Word reading accuracy.** The Word Identification subtest of the Woodcock Reading Mastery Tests – Revised (Form G) (Woodcock, 1998) measures context-free word reading ability. Participants were tested individually and were asked to read words of increasing difficulty aloud. Testing was ended after six consecutive errors. The score is the total number of words read correctly. Scores could range from 0 to 106, with higher scores indicating greater ability. The split-half reliability estimate with Spearman-Brown correction in this sample of .81 is similar to .86 reported for this age group in the test manual.

**Phonological decoding.** The Word Attack subtest from the Woodcock Reading Mastery Tests – Revised (Form G) (Woodcock, 1998) assesses participants’ ability to apply phonic and structural analysis skills in pronouncing pseudowords (e.g., *dee, monglustamer*). This test contains 45 items arranged in order of increasing difficulty. For a response to be scored correct, the participant must produce a natural reading of the word in about 5 seconds. Testing is ended after six consecutive errors. The score is the number of pseudowords read correctly. Scores could range from 0 to 45 with higher scores indicating greater ability. The split-half coefficient with
Spearman-Brown correction in this sample was .76, compared to .84 reported for this age group in the test manual.

**Word reading rate.** The Test of Word Reading Efficiency (Form A) (Torgesen, Wagner, & Rashotte, 1999) measures the ability of participants to pronounce printed words accurately and fluently. Participants have 45 seconds to read aloud as many words (maximum 104) that are arranged in increasing order of difficulty. The score is the number of words read correctly in 45 sec. Scores could range from 0 to 104 with higher scores indicating greater ability. The test manual gives an alternative-form reliability coefficient for 15 year olds of .88; I could not calculate the equivalent score because I did not give an alternative form of the test.

**Nonverbal ability.** The Matrix Analogies Test (MAT-SF; Naglieri, 1985) is a measure of fluid intelligence. It is comprised of 1 practice item and 34 test items, each of which shows an incomplete pattern and six answer alternatives. Following the manual’s instructions, the test was administered in group format with a 25-minute time limit. One mark was awarded for each correct response. Scores could range from 0 to 34 with higher scores indicating greater nonverbal ability. The Cronbach’s alpha value for this sample was .85, similar to the .83 median Cronbach’s alpha (5- to 17-year age group) reported in the test manual.

**Vocabulary.** Test-V of the Canadian Tests of Basic Skills – English (Nelson, 1998) (Form K, Level 16) is a 40-item measure of receptive vocabulary skill in which participants read a target word contained in a phrase (e.g., “realized they have been duped”), and select a synonym from five other words (e.g., discovered, fooled, replaced, beaten, forgotten). Following the instructions prescribed in the manual, the test was administered in group format with a 15 minute time limit. Each correct response scores 1 point. Scores could range from 0 to 40 with higher
scores indicating greater vocabulary knowledge. The split half reliability coefficient was .84, similar to .86 reported in the test manual (Nelson, 1998).

**Individual differences measures.** Measures of print exposure, working memory, and approaches to learning were administered.

*Print exposure.* An adapted version of the *Author Recognition Test* (Stanovich, West & Harrison, 1995) was used to measure participants’ print exposure (see Appendix B). It is an untimed, group administered recognition checklist (50 real authors and 50 foils). Because the names of the authors included in the original measure were designed to assess the print exposure of college-aged readers, thirty percent of the names were replaced with the names of authors considered appropriate for high school students. Each participant’s score was the number of correct responses minus the number of incorrect responses. Higher scores indicated greater exposure to print. Cronbach’s alpha was .92.

*Working memory.* An individually administered backward digit span task (see Appendix C) was used to assess participants’ ability to simultaneous store and process verbal information. Participants were asked to repeat, in reverse order, orally presented digits. Six trials were presented at each series length (2 to 9). A stop rule of 4 incorrect out of the last 6 items was used. Each correct trial scores 1 point. Scores could range from 0 to 48 with higher scores indicating higher working memory. The split-half reliability was .94.

*Student approaches to learning.* The Revised Two-Factor Study Process Questionnaire (Biggs, Kember & Leung, 2001) is a self-report questionnaire that assesses students’ usual way of studying. Students respond to 20 Likert scale items. The test consists of two scales, one measuring the deep approach to learning (e.g., *I find that at times studying gives me a feeling of*...
deep personal satisfaction.), and the other measuring the surface approach (e.g., I do not find my school work very interesting so I keep my work to the minimum.). Students respond on a 5-point scale. Scores for each approach can range from 10 to 50 with higher scores indicating greater use of that approach. Responses were missing for three subjects (total 4 items), and these missing values were replaced by the sample means. The measure was group administered without a time limit. The Cronbach’s alpha values for the Deep Approach and the Surface Approach are similar to those reported by Kember, Biggs, and Leung (2004) (.81 and .69 versus .82 and .71 respectively).

**Word-related measures.** Measures of phonological awareness, orthographic knowledge, morphological awareness were administered.

**Phonological awareness.** The Elision subtest of the *Comprehensive Test of Phonological Processing* (Wagner, Torgesen, & Rashotte, 1999) is an individually administered test that assesses students’ ability to delete specified sounds from words. Students repeat a word, and then say the word formed after deleting a specified sound (e.g., say the word cup; say the word cup without the /k/; correct answer: up). There are three practice items and 20 test items. Testing is discontinued after three consecutive errors. One point is given for each correct response. Scores could range from 0 to 20, with higher scores indicating greater ability. Cronbach’s alpha was .73.

**Orthographic knowledge.** This measure was constructed from items selected from Olson, Forsberg, Wise, and Rack (1994). In this timed group administered test, students circled the correctly spelled words from 30 pairs of irregular words and their pseudohomophones (e.g., answer – anser). Performance was measured by the number of items correctly marked minus the
number of incorrectly marked items within 45 seconds. The split half reliability coefficient was .84.

*Morphological awareness.* An adapted version of the Morphological Base Identification Circling Task (Bowers & Kirby, 2010) was used to assess participants’ ability to identify the bases of a variety of multi-morphemic words, presented in written form. Because the task had originally been carried out with a younger group, the adaptation included an additional 10 words that were more challenging. There were 40 items in this untimed, group administered test. Participants were instructed to circle “the main part” of each word. Five practice words were presented. The scoring of this task was on a 3-point scale: 2 points for circling the base or smallest stem (a base with at least one affix) of a word that could stand on its own as a word; 1 point for circling any part of a word that removed at least one affix, but failed to reduce the target word to the smallest stem that could stand on its own as a word; and 0 points otherwise. Scores could range from 0 to 80, with higher scores indicating greater ability. Cronbach’s alpha was .73.

**Reading comprehension.** Four measures of reading comprehension were administered.

*Canadian Test of Basic Skills.* Test-R (Reading) of the Canadian Tests of Basic Skills (Nelson, 1998) – English (Form K, Level 16) is adapted from Form K of the Iowa Test of Basic Skills. The test assesses the literal and inferential understanding of reading passages. Following the instructions outlined in the publisher’s manual, the test was group administered with 40 minutes to answer 44 questions based on five reading passages (two narratives, one poem, and two expositions). Multiple-choice questions follow each passage with five answer choices per item. Each correct response scores 1 point for a maximum of 44. The split-half coefficient of .83 in this sample is similar to the KR-20 value of .89 reported in the manual.
Novel word meaning. This untimed, group administered task required students to read 3 narrative and 3 expository experimental passages (117-165 words each). Each passage contained a novel word (e.g., slond, Diboll). Students were asked to use the information contained in the text to infer the novel word’s meaning, which they did in writing. The number of pieces of information and the distance between the novel word and its defining information varied between texts (3 - 4 pieces of defining information, distributed within and across sentence boundaries). Each text scored a maximum of 3 or 4 points, according to the number pieces of defining information. The maximum possible score was 20. Fifteen of the responses were scored individually by the two authors and results compared; inter-rater reliability was .86. Any disagreements were resolved through discussion, and the remaining summaries were scored by the first author. Cronbach’s alpha was .69. Passages and the scoring scheme are included in Appendix D.

Syllogisms. This was an untimed, group administered deductive reasoning task that was adapted from Stanovich, West and Harrison (1995). Each item required students to determine the validity of the conclusion, given two premises. Because some of the premises were counterfactual, the task required participants to suppress their own background knowledge, and instead rely on the information contained in the two premises of the text. There were 23 items each consisting of two premises and a conclusion. There were three types of items (counterfactual, consistent factual, and neutral). Eight of the items had a counterfactual conclusion that was either logically valid or invalid from the premises (4 items each); for example, All things that are smoked are good for your health (A); Cigarettes are smoked (B); Cigarettes are good for your health (C); valid. Eight items were factually consistent with prior
knowledge, being either logically valid or invalid (4 items each); for example, *All insects need oxygen (A); Mice need oxygen (B); Mice are insects (C)*; invalid. Seven items involved imaginary content and were thus neutral with respect to prior knowledge; these were either logically valid or invalid (4 and 3 each respectively); for example, *All opprobines run on electricity (A); Jamtops run on electricity (B); Jamtops are opprobines (C)*; invalid. One point was given for each correct response and one point was subtracted for each incorrect response. Cronbach’s alpha was .68.

**Summary writing.** This task asked students to write a text-absent summary of an expository text. Before this task began, students were reminded what constituted a good summary, and informed that the text and any notes they made while reading would be removed before they wrote their summaries. They were then given 12 minutes to read, study, and take notes on a 680 word expository passage titled Controlling Pests (taken from Manning, 1989) that discussed the ecological and environmental problems associated with the use of pesticides and natural-enemy pest control. After a 5 minute interval, in which the original text and notes taken by the students were collected, they were given 10 minutes to write a text-absent summary. Scoring criteria were modelled after those used by Manning (1989), Kirby and Pedwell (1991), Kirby and Woodhouse (1994), and Stein and Kirby (1992). The text was analyzed into propositions (idea units) at three levels of depth: (a) *details* required little integration across text units and were stated relatively explicitly in the text; these were divided into important and unimportant details, depending on their relation to the main ideas, and only the important ones were scored; (b) *main ideas* required the reader to integrate details included in the text, but did not require the reader to construct an overall understanding or abstraction of the entire text; and (c) *themes* required the reader to
construct an integrated and abstracted mental representation of the text, using processes such as inferencing, elaborating, integrating, and relating for the generation and comprehension of this abstract, unstated information (Kirby & Cantwell, 1985). The first two categories were intended to reflect the microstructure and macrostructure levels described by Kintsch (1998); the theme category was intended to address content that could only be found in the situation model. Using this method, a scoring template of the text was developed (24 important detail units, 12 main idea units, and 4 thematic units), and then applied to the summaries. Twenty of the summaries were scored by the two authors and results compared; inter-rater reliability was .95. Any disagreements were resolved through discussion, and the remaining summaries were scored by the first author. The number of words written in each summary were also counted. The original passage and scoring scheme are included in Appendices E and F, respectively.

**Results**

**Descriptive Statistics**

Table 3.1 reports descriptive statistics for all measures for the total sample of 137. Following the guidelines of Tabachnick and Fidell (2007), univariate normality for all variables was assessed: univariate outliers (z-scores greater than or less than 3 standard deviations) and skewness problems (skewness coefficient/standard error greater than or less than 3) were identified. Three cases of outlying low performance were found, all for phonological awareness; the scores were changed to one less than the lowest score obtained by the remaining subjects. Five variables (word reading accuracy, phonological decoding, nonverbal ability, phonological awareness, and novel words) were negatively skewed, and two variables (working memory and exposure to print) were positively skewed. Square root transformations, with reflection when necessary, corrected the skewness problems. Subsequent analyses were conducted with the
original and with the square-root-transformed data, but no differences in interpretation resulted; therefore, I report the analyses using the untransformed scores. The correlations among all variables are reported in Table 3.2.
<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word reading accuracy</td>
<td>87.1</td>
<td>8.1</td>
</tr>
<tr>
<td>Phonological decoding</td>
<td>34.4</td>
<td>6.1</td>
</tr>
<tr>
<td>Word reading rate</td>
<td>86.3</td>
<td>10.4</td>
</tr>
<tr>
<td>Nonverbal ability</td>
<td>28.3</td>
<td>4.0</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>18.0</td>
<td>7.2</td>
</tr>
<tr>
<td><strong>Individual Differences Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure to print</td>
<td>6.1</td>
<td>3.9</td>
</tr>
<tr>
<td>Working memory</td>
<td>70.1</td>
<td>36.4</td>
</tr>
<tr>
<td>Deep Approach</td>
<td>22.4</td>
<td>6.1</td>
</tr>
<tr>
<td>Surface Approach</td>
<td>29.2</td>
<td>6.1</td>
</tr>
<tr>
<td><strong>Word-related Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonological awareness</td>
<td>17.1</td>
<td>2.3</td>
</tr>
<tr>
<td>Orthographic knowledge</td>
<td>37.1</td>
<td>6.9</td>
</tr>
<tr>
<td>Morphological awareness</td>
<td>43.1</td>
<td>7.3</td>
</tr>
<tr>
<td><strong>Reading comprehension Measures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canadian Test of Basic Skills</td>
<td>24.8</td>
<td>8.9</td>
</tr>
<tr>
<td>Novel words</td>
<td>14.4</td>
<td>3.5</td>
</tr>
<tr>
<td>Syllogisms</td>
<td>7.8</td>
<td>6.5</td>
</tr>
<tr>
<td>Summary writing important details</td>
<td>7.3</td>
<td>4.7</td>
</tr>
<tr>
<td>Summary writing main ideas</td>
<td>5.1</td>
<td>2.4</td>
</tr>
<tr>
<td>Summary writing themes</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Summary writing number of words</td>
<td>125.7</td>
<td>41.3</td>
</tr>
</tbody>
</table>
Table 3.2

**Correlations Among all Variables (N = 137)**

<table>
<thead>
<tr>
<th></th>
<th>WRA</th>
<th>PD</th>
<th>WRR</th>
<th>NVA</th>
<th>VOC</th>
<th>EXP</th>
<th>WM</th>
<th>DA</th>
<th>SA</th>
<th>PA</th>
<th>ORT</th>
<th>MOR</th>
<th>CTB</th>
<th>NW</th>
<th>SYL</th>
<th>SD</th>
<th>SM</th>
<th>ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>WRA</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD</td>
<td>.72**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WRR</td>
<td>.66**</td>
<td>.60**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVA</td>
<td>.34**</td>
<td>.12</td>
<td>.23**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VOC</td>
<td>.58**</td>
<td>.33**</td>
<td>.49**</td>
<td>.41**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXP</td>
<td>.49**</td>
<td>.32**</td>
<td>.42**</td>
<td>.15†</td>
<td>.61**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WM</td>
<td>.39**</td>
<td>.32**</td>
<td>.47**</td>
<td>.28**</td>
<td>.42**</td>
<td>.38**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DA</td>
<td>.03</td>
<td>.04</td>
<td>.16†</td>
<td>.11</td>
<td>.11</td>
<td>.01</td>
<td>.18*</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SA</td>
<td>-.13</td>
<td>-.12</td>
<td>-.11</td>
<td>-.17</td>
<td>-.12</td>
<td>-.05</td>
<td>-.12</td>
<td>-.35**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>.42**</td>
<td>.40**</td>
<td>.41**</td>
<td>.15†</td>
<td>.29**</td>
<td>.18*</td>
<td>.35**</td>
<td>-.01</td>
<td>.01</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORT</td>
<td>.47**</td>
<td>.31**</td>
<td>.62**</td>
<td>.13</td>
<td>.24**</td>
<td>.26**</td>
<td>.29**</td>
<td>.24**</td>
<td>-.07</td>
<td>.21*</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOR</td>
<td>.38**</td>
<td>.28**</td>
<td>.23**</td>
<td>.31**</td>
<td>.27**</td>
<td>.22*</td>
<td>.22*</td>
<td>-.08</td>
<td>-.06</td>
<td>.20*</td>
<td>.25**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTB</td>
<td>.56**</td>
<td>.32**</td>
<td>.56**</td>
<td>.36**</td>
<td>.66**</td>
<td>.45**</td>
<td>.45**</td>
<td>.21*</td>
<td>.17*</td>
<td>.19*</td>
<td>.45**</td>
<td>.37**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NW</td>
<td>.48**</td>
<td>.23**</td>
<td>.35**</td>
<td>.41**</td>
<td>.49**</td>
<td>.20*</td>
<td>.29**</td>
<td>.15†</td>
<td>-.06</td>
<td>.28**</td>
<td>.29**</td>
<td>.32**</td>
<td>.54**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SYL</td>
<td>.32**</td>
<td>.30**</td>
<td>.38**</td>
<td>.32**</td>
<td>.46**</td>
<td>.26**</td>
<td>.38**</td>
<td>.07</td>
<td>-.08</td>
<td>.16†</td>
<td>.15†</td>
<td>.15†</td>
<td>.44**</td>
<td>.34**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>.27**</td>
<td>.17*</td>
<td>.25**</td>
<td>.29**</td>
<td>.32**</td>
<td>.16†</td>
<td>.32**</td>
<td>.11</td>
<td>.02</td>
<td>.29**</td>
<td>.24**</td>
<td>.28**</td>
<td>.37**</td>
<td>.28**</td>
<td>.08</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SM</td>
<td>.31**</td>
<td>.28**</td>
<td>.38**</td>
<td>.23**</td>
<td>.35**</td>
<td>.13</td>
<td>.33**</td>
<td>.13</td>
<td>-.03</td>
<td>.33**</td>
<td>.32**</td>
<td>.21*</td>
<td>.44**</td>
<td>.40**</td>
<td>.13</td>
<td>.68**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>ST</td>
<td>.18*</td>
<td>.11</td>
<td>.17*</td>
<td>-.01</td>
<td>.22**</td>
<td>.21*</td>
<td>.10</td>
<td>.22**</td>
<td>-.11</td>
<td>-.02</td>
<td>-.04</td>
<td>.29**</td>
<td>.26**</td>
<td>.20*</td>
<td>.23**</td>
<td>.03</td>
<td>-.23</td>
<td>.03</td>
</tr>
<tr>
<td>SW</td>
<td>.30**</td>
<td>.19*</td>
<td>.38**</td>
<td>.25**</td>
<td>.28**</td>
<td>.17*</td>
<td>.29**</td>
<td>.16†</td>
<td>.05</td>
<td>.25**</td>
<td>.34**</td>
<td>.26**</td>
<td>.39**</td>
<td>.42**</td>
<td>.17*</td>
<td>.70**</td>
<td>.64**</td>
<td>.08</td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, †p < .08. WRA=Word reading accuracy, PD=Phonological decoding, WRR=Word reading rate, NVA=Non-verbal ability, VOC=Vocabulary, EXP=Exposure to Print, WM=Working memory, DA=Deep Approach, SA=Surface Approach, PA=Phonological awareness, ORT=Orthographic Knowledge, MOR=Morphological Awareness, CTB=Canadian Test of Basic Skills, NW=Novel words, SYL=Syllogisms, SD=Summary writing details, SM=Summary writing main ideas, ST=Summary writing themes, SW=Summary writing words written.
Models Using Different Control Variables

In the analyses that follow, I systematically examine the effects of using five control variables to identify the unexpected poor comprehenders: word reading accuracy, phonological decoding, word reading rate, nonverbal ability, and vocabulary. I do this to determine whether (a) the comprehension weaknesses of unexpected poor comprehenders are associated with concurrent weaknesses in these variables when they are not controlled, (b) differences in the mean scores of unexpected poor comprehenders for those variables that are not controlled depend upon the contrast group (expected average or unexpected good comprehenders), and (c) these different control procedures identify different groups of unexpected poor comprehenders.

Using the total sample (N=137), I regressed reading comprehension (defined as the Canadian Test of Basic Skills score) on different combinations of the control variables. In each of these analyses, the predicted values for the outcome were graphed against the actual reading comprehension scores (see Figure 3.1 for an example). Based on the regression line and confidence intervals about it, individuals were selected for three groups: unexpected poor comprehenders, expected average comprehenders, and unexpected good comprehenders. To make the groups more distinct, buffer zones were established between groups: individuals whose actual values were between the 40th and 60th percentile confidence intervals of the regression line were excluded. Unexpected poor comprehenders had actual values that fell below the 60th percentile confidence interval; thus their reading comprehension was less than would be predicted from the control variables. Unexpected good comprehenders’ actual values fell above the 60th percentile confidence interval; thus their reading comprehension scores were greater than would be expected from the control variables. Finally, expected average comprehenders’ actual
values fell within the 40th percentile confidence intervals, indicating that their reading comprehension scores were consistent with the control variables. Furthermore, to exclude students with weak basic skills (to meet the fundamental definition of unexpected poor comprehenders), a minimum criterion of -1 on the standardized predicted value (i.e., one standard deviation below the mean) was established.

Figure 3.1

*Scatter Plot for Selection of Unexpected Poor, Expected Average, and Unexpected Good Comprehenders, in Model 4 (word reading accuracy, reading rate, nonverbal ability, and vocabulary)*

*Note.* Some symbols overlap with one another.
Five regression models were calculated, with different sets of predictors. Because it had been used in most previous studies, word reading accuracy was in every model. The predictors in each of the models were as follows: model 1: word reading accuracy; model 2: word reading accuracy and reading rate; model 3: word reading accuracy and nonverbal ability; model 4: word reading, reading rate, and nonverbal ability; and model 5: word reading, reading rate, nonverbal ability and vocabulary. Because no significant between-group differences were observed for the phonological decoding variable in any of these models, its inclusion as a predictor in regression analyses was not necessary; presumably any effect of it was included in word reading accuracy. When vocabulary was added in the final regression model, significant between-group differences occurred for some of the other predictor variables; to avoid eliminating more participants to match groups, I decided to use vocabulary as a covariate in later analyses. I therefore report results for four models. For each group in each model, I also examined individual profiles for each of the predictor variables entered into the regression equations to determine whether performance on these variables reflected group characteristics (e.g., z-score performance on word reading rate for unexpected poor comprehenders was less than z-score for reading comprehension). Although there were some odd cases, I argue that the point of the regression approach is to set up the overall model, with the combined prediction being the real prediction. Individual scores on individual tests contain error, whereas the predicted score should have error removed.

**Performance on control measures.** Table 3.3 reports the means and standard deviations of the three groups for the five control variables and the reading comprehension outcome.
### Table 3.3

**Means and Standard Deviations of Unexpected Poor, Expected Average, and Unexpected Good Comprehenders on Predictor and Outcome Variables in Four Models**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>WRA</td>
<td>WRA + WRR</td>
<td>WRA + NVA</td>
<td>WRA + WRR + NVA</td>
</tr>
<tr>
<td>Word reading accuracy (WRA)</td>
<td>Unexpected Poor</td>
<td>88.5</td>
<td>88.3</td>
<td>89.2</td>
<td>88.4</td>
</tr>
<tr>
<td></td>
<td>Expected Average</td>
<td>89.3</td>
<td>89.4</td>
<td>89.0</td>
<td>87.9</td>
</tr>
<tr>
<td></td>
<td>Unexpected Good</td>
<td>90.6</td>
<td>90.0</td>
<td>91.0</td>
<td>90.3</td>
</tr>
<tr>
<td>Phonological decoding</td>
<td>Unexpected Poor</td>
<td>36.0</td>
<td>36.6</td>
<td>36.1</td>
<td>36.4</td>
</tr>
<tr>
<td></td>
<td>Expected Average</td>
<td>35.2</td>
<td>36.0</td>
<td>34.6</td>
<td>34.7</td>
</tr>
<tr>
<td></td>
<td>Unexpected Good</td>
<td>36.2</td>
<td>35.8</td>
<td>36.3</td>
<td>35.6</td>
</tr>
<tr>
<td>Word reading rate (WRR)</td>
<td>Unexpected Poor</td>
<td>84.2</td>
<td>87.9</td>
<td>83.5</td>
<td>87.1</td>
</tr>
<tr>
<td></td>
<td>Expected Average</td>
<td>87.1</td>
<td>89.5</td>
<td>86.4</td>
<td>88.0</td>
</tr>
<tr>
<td></td>
<td>Unexpected Good</td>
<td>92.3</td>
<td>90.3</td>
<td>92.5</td>
<td>90.4</td>
</tr>
<tr>
<td>Nonverbal ability (NVA)</td>
<td>Unexpected Poor</td>
<td>26.9</td>
<td>25.9</td>
<td>28.4</td>
<td>28.4</td>
</tr>
<tr>
<td></td>
<td>Expected Average</td>
<td>28.5</td>
<td>28.3</td>
<td>29.5</td>
<td>29.1</td>
</tr>
<tr>
<td></td>
<td>Unexpected Good</td>
<td>30.5</td>
<td>30.3</td>
<td>28.7</td>
<td>29.4</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>Unexpected Poor</td>
<td>14.6</td>
<td>14.3</td>
<td>16.1</td>
<td>16.0</td>
</tr>
</tbody>
</table>
measure used in the regression analyses for each of the four models; as shown, groups ns varied slightly from model to model. ANOVA results and post hoc contrasts (t-tests with Bonferroni correction) are shown in Table 3.4. Overall, the groups differed on variables that were not controlled but did not differ when those variables were controlled (or on Phonological decoding). When there were group differences, unexpected poor and unexpected good comprehenders always differed; the performance of the expected average comprehenders was intermediate, often lower than that of the unexpected good comprehenders but not different from that of the unexpected poor comprehenders. Table 3.3 also shows the reading comprehension scores of the three groups in the four models. As expected, all groups differed in all models.

<table>
<thead>
<tr>
<th>Reading Comprehension</th>
<th>Expected Average</th>
<th>Unexpected Good</th>
<th>Unexpected Poor</th>
<th>Expected Average</th>
<th>Unexpected Good</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17.7 5.9</td>
<td>25.2 6.1</td>
<td>15.1 3.6</td>
<td>26.6 4.0</td>
<td>36.1 3.4</td>
</tr>
<tr>
<td></td>
<td>18.1 6.2</td>
<td>24.1 6.7</td>
<td>14.6 3.7</td>
<td>27.1 4.7</td>
<td>35.6 3.4</td>
</tr>
<tr>
<td></td>
<td>17.3 5.1</td>
<td>24.3 7.1</td>
<td>15.4 3.9</td>
<td>27.1 4.0</td>
<td>36.0 3.5</td>
</tr>
<tr>
<td></td>
<td>17.1 5.2</td>
<td>24.9 6.6</td>
<td>15.5 4.2</td>
<td>26.7 4.4</td>
<td>35.9 3.4</td>
</tr>
</tbody>
</table>

*Note. Group sizes for Unexpected Poor, Expected Average, and Unexpected Good Comprehenders, respectively: Model 1, 31, 38, 25; Model 2, 24, 38, 24; Model 3, 27, 35, 24; Model 4, 28, 37, 21.*
Table 3.4  
Summary of Analyses of Variance Comparing Unexpected Poor, Expected Average, and Unexpected Good Comprehenders on Control Variables for Four Models

<table>
<thead>
<tr>
<th>Control Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WRA</td>
<td>WRA + WRR</td>
<td>WRA + NVA</td>
<td>WRA + WRR + NVA</td>
</tr>
<tr>
<td>Word Reading Accuracy (WRA)</td>
<td>F(2,91)</td>
<td>F(2,83)</td>
<td>F(2,83)</td>
<td>F(2,83)</td>
</tr>
<tr>
<td></td>
<td>1.07</td>
<td>.61</td>
<td>1.25</td>
<td>1.29</td>
</tr>
<tr>
<td>Phonological Decoding</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>.53</td>
<td>.31</td>
<td>1.43</td>
<td>1.20</td>
</tr>
<tr>
<td>Word reading rate (WRR)</td>
<td>5.41**</td>
<td>.55</td>
<td>6.00**</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td>P&lt; G**</td>
<td>-</td>
<td>P&lt; G**</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>A&lt; G†</td>
<td></td>
<td>A&lt; G*</td>
<td></td>
</tr>
<tr>
<td>Nonverbal ability (NVA)</td>
<td>5.53**</td>
<td>8.46***</td>
<td>.79</td>
<td>.65</td>
</tr>
<tr>
<td></td>
<td>P&lt; G**</td>
<td>P&lt; G***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>P&lt; A*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>21.8***</td>
<td>16.50***</td>
<td>14.68***</td>
<td>18.32***</td>
</tr>
<tr>
<td></td>
<td>P&lt; G***</td>
<td>P&lt; G***</td>
<td>P&lt; G***</td>
<td>P&lt; G***</td>
</tr>
<tr>
<td></td>
<td>A&lt; G***</td>
<td>A&lt; G**</td>
<td>A&lt; G***</td>
<td>A&lt; G***</td>
</tr>
<tr>
<td></td>
<td>P&lt; A*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>224.20***</td>
<td>160.39***</td>
<td>188.98***</td>
<td>150.41***</td>
</tr>
<tr>
<td></td>
<td>P&lt; G***</td>
<td>P&lt; G***</td>
<td>P&lt; G***</td>
<td>P&lt; G***</td>
</tr>
<tr>
<td></td>
<td>A&lt; G***</td>
<td>A&lt; G***</td>
<td>A&lt; G***</td>
<td>A&lt; G***</td>
</tr>
<tr>
<td></td>
<td>P&lt; A***</td>
<td>P&lt; A***</td>
<td>P&lt; A***</td>
<td>P&lt; A***</td>
</tr>
</tbody>
</table>

*Note.* P unexpected poor comprehenders, A expected average comprehenders, G unexpected good comprehenders. *p < .05, ** < .01, *** < .001; † p < .10
**Consistency of identification across models.** I next addressed the extent to which the same individuals were identified in each of the groups across the four models. Table 3.5 reports the percentage of students in each model according to the groups in which they appeared in other models. Considerable overlap was evident across the four models, ranging from 65% to 95%. Furthermore, a substantial number of the same individuals were identified into the same skill group in each of the four models: 17 unexpected poor comprehenders, representing 55% to 71% of this group; 21 expected average comprehenders (55% to 60%); and 16 unexpected good comprehenders (64% to 76%). I also examined the extent to which individuals assigned into one group were assigned to a different group under a different model. There was very little evidence for group change across models: 3 unexpected poor comprehenders in one model were identified as expected average comprehenders in another model, and 1 unexpected good comprehender became an expected average comprehender in another model. In all other cases, individuals identified in a specific group in one model, but not in another, were in the buffer zone.
Table 3.5

Consistency of Group Membership across Four Models (rows show percent of students classified under each model of those classified in column model)

<table>
<thead>
<tr>
<th>Model</th>
<th>Control Variables</th>
<th>Groups</th>
<th>n</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Word Reading Accuracy</td>
<td>Poor</td>
<td>31</td>
<td>-</td>
<td>71</td>
<td>81</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>38</td>
<td>74</td>
<td>76</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Good</td>
<td>25</td>
<td>84</td>
<td>76</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Word Reading Accuracy + Word Reading Rate</td>
<td>Poor</td>
<td>24</td>
<td>92</td>
<td>-</td>
<td>75</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>38</td>
<td>74</td>
<td>68</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Good</td>
<td>24</td>
<td>88</td>
<td>83</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Word Reading Accuracy + Nonverbal Ability</td>
<td>Poor</td>
<td>27</td>
<td>93</td>
<td>70</td>
<td>-</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>35</td>
<td>80</td>
<td>71</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Good</td>
<td>24</td>
<td>83</td>
<td>83</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Word Reading Accuracy + Word Reading Rate + Nonverbal Ability</td>
<td>Poor</td>
<td>28</td>
<td>82</td>
<td>71</td>
<td>75</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average</td>
<td>37</td>
<td>65</td>
<td>81</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Good</td>
<td>21</td>
<td>86</td>
<td>95</td>
<td>95</td>
<td></td>
</tr>
</tbody>
</table>
Because of the group differences that remained when fewer control variables were used (see Table 3.4) it was decided to use Model 4, in which word reading accuracy, word reading rate, and nonverbal ability were the predictors, for further analyses. Because Vocabulary could not be controlled in the selection process, it was used as a covariate in the subsequent analyses.

**Differences between Comprehension Groups**

Performance of the unexpected poor, expected average, and unexpected good comprehender groups identified in Model 4 was compared for three sets of variables: individual differences (working memory, print exposure, and approaches to learning), word-related measures (phonological awareness, orthographic knowledge, and morphological awareness), and comprehension-related measures (novel words; syllogisms; details, main ideas, themes, and number of words written from the summary writing task). Table 3.6 presents the means and standard deviations of each of these measures for each group.
Table 3.6

Means and Standard Deviations for Individual Differences, Word Related, and Reading Comprehension Variables of Unexpected Poor, Expected Average, and Unexpected Good Comprehenders (Model 4)

<table>
<thead>
<tr>
<th></th>
<th>Unexpected Poor Comprehenders</th>
<th>Expected Average Comprehenders</th>
<th>Unexpected Good Comprehenders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td><strong>Individual Difference Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working Memory</td>
<td>63.3</td>
<td>20.7</td>
<td>71.7</td>
</tr>
<tr>
<td>Exposure to Print</td>
<td>5.9</td>
<td>3.9</td>
<td>5.4</td>
</tr>
<tr>
<td>Deep Approach</td>
<td>20.1</td>
<td>5.8</td>
<td>22.2</td>
</tr>
<tr>
<td>Surface Approach</td>
<td>31.4</td>
<td>4.4</td>
<td>29.5</td>
</tr>
<tr>
<td><strong>Word-Related Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonological Awareness</td>
<td>17.4</td>
<td>2.1</td>
<td>17.2</td>
</tr>
<tr>
<td>Orthographic Knowledge</td>
<td>35.5</td>
<td>7.2</td>
<td>37.9</td>
</tr>
<tr>
<td>Morphological Awareness</td>
<td>42.8</td>
<td>6.5</td>
<td>42.7</td>
</tr>
<tr>
<td><strong>Comprehension-related Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel words</td>
<td>13.4</td>
<td>3.2</td>
<td>14.3</td>
</tr>
<tr>
<td>Syllogisms</td>
<td>6.5</td>
<td>5.4</td>
<td>7.2</td>
</tr>
<tr>
<td>Summary Details</td>
<td>5.7</td>
<td>3.7</td>
<td>8.8</td>
</tr>
<tr>
<td>Summary Main Ideas</td>
<td>4.3</td>
<td>1.9</td>
<td>5.8</td>
</tr>
<tr>
<td>Summary Theme</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Summary Words Written</td>
<td>115.1</td>
<td>35.3</td>
<td>133.2</td>
</tr>
</tbody>
</table>
Three multivariate analyses of variance (MANOVAs) were carried out, one for each set of variables (individual differences, word-related, and comprehension-related). Multivariate analyses of covariance (MANCOVAs) were also carried out, with Vocabulary as the covariate. Using the guidelines provided by Stevens (2001), I examined whether multivariate assumptions were tenable. Each variable for each group was examined by identifying influential outliers (i.e., $z$-scores greater than or less than 3). Three outliers were identified. One unexpected good comprehender had an extreme high value on morphological awareness; this score was replaced with a new score of 1 more than the remaining highest value. One unexpected poor comprehender had an extreme low score on phonological awareness; this was replaced with a score of 1 less than the remaining lowest value. One unexpected good comprehender had an extreme low score on novel words; this was replaced with a score 1 less than the next lowest score. Given the difficulty of meeting the assumptions of multivariate normality, homogeneity of variance, and homogeneity of covariance matrices, I follow Stevens (2001) that when group sizes are approximately equal MANOVA is robust to Type I and II errors. When MANOVA or MANCOVA test statistics were significant, univariate effects (ANOVA and ANCOVA) are reported. Follow-up pairwise group comparisons were conducted when the ANOVAs or ANCOVAs were significant; Bonferroni correction is reported in cases when homogeneity of variances between pairs of variables could be assumed (Levene’s $p > .05$), Tamhane correction when homogeneity of variance could not be assumed (Levene’s $p < .05$).

**Individual differences measures.** The MANOVA for the individual differences variables (working memory, exposure to print, deep approach, surface approach) indicated a significant group effect, Wilks’ $\Lambda = .78$, $F(8,160) = 2.59$, $p = .01$. Post hoc ANOVAs indicated
significant group effects for working memory, $F(2, 83) = 3.74, p = .03, \eta^2 = .08$, and exposure to print, $F = 6.34, p = .003, \eta^2 = .13$, a marginal effect for surface approach, $F(2, 83) = 2.72, p = .07, \eta^2 = .06$, and a nonsignificant effect for deep approach, $F(2, 83) = 1.93, p = .15, \eta^2 = .05$. Pairwise comparisons revealed that the unexpected poor comprehenders were significantly lower than the unexpected good comprehenders for exposure to print ($p = .02$, Bonferroni), marginally lower on working memory ($p = .08$, Tamhane), and marginally higher for surface approach ($p = .06$, Tamhane). No significant differences were found between unexpected poor and expected average comprehenders for any of the individual differences variables (all $p$s > .38). Expected average comprehenders and unexpected good comprehenders differed only in print exposure ($p = .016$ Tamhane), with the unexpected good comprehenders having a substantial advantage. In the MANCOVA, there was a significant effect for vocabulary, Wilks’ $\Lambda = .73, F(4,79) = 7.38, p < .001$, but no significant effect of group, Wilks’ $\Lambda = .90, F(8,158) = 1.09, p = .38$, so no further analyses were performed.

**Word-related measures.** The MANOVA for the word related measures (phonological awareness, orthographic knowledge, morphological awareness) revealed a significant group effect, Wilks’ $\Lambda = .856, F(6,162) = 2.18, p = .047$. Post hoc ANOVAs indicated marginal group effects for orthographic knowledge, $F(2, 83) = 2.89, p = .061, \eta^2 = .07$, and morphological awareness, $F(2, 83) = 2.85, p = .064, \eta^2 = .06$, but not for phonological awareness, $F(2, 83) = .81, p = .45, \eta^2 = .02$. To inform future research, I followed-up the marginally significant effects with pairwise comparisons. For orthographic knowledge, the unexpected poor comprehenders performed significantly worse than the unexpected good comprehenders ($p = .04$, Tamhane), but did not differ significantly from the expected average comprehenders ($p = .13$, Bonferroni). For
morphological awareness, the unexpected poor comprehenders and expected average comprehenders did not differ \( (p = .99, \text{ Bonferroni}) \), but both of these groups performed marginally worse than the unexpected good comprehenders \( (p = .08, .09, \text{ Tamhane, respectively}) \). For the MANCOVA, no significant effects were found for vocabulary, Wilks’ \( \Lambda = .936, F (3,80) = 1.82, p = .15 \), or group, Wilks’ \( \Lambda = .911, F (6,160) = 1.28, p = .27 \), so no post hoc analyses were done.

**Comprehension-related measures.** The MANOVA for the reading comprehension measures was significant, Wilks’ \( \Lambda = .58, F (12,156) = 12.00, p < .001 \). Results for post hoc ANOVA and pairwise comparisons are reported in Table 3.7. Univariate ANOVA results revealed significant group effects for all reading comprehension variables. Pairwise comparisons indicated that the unexpected poor comprehenders always performed worse than the unexpected good comprehenders, and that the unexpected poor comprehenders performed worse than the expected average comprehenders on the detail and main idea summary scores. In the MANCOVA, the vocabulary effect was marginally significant, Wilks’ \( \Lambda = .87, F (6,77) = 1.99, p = .08 \), and the group effect remained significant, Wilks’ \( \Lambda = .73, F (12,154) = 2.18, p = .015 \). Post hoc ANCOVAs and pairwise comparisons showed significant group effects for all of the summary variables except number of words written; the unexpected poor comprehenders performed worse than the expected average comprehenders on summary details and main ideas, and the latter group performed worse than the unexpected good comprehenders on summary themes.
Table 3.7

ANOVA and ANCOVA Results for Comprehension-related Variables (Model 4)

<table>
<thead>
<tr>
<th>MANOVA</th>
<th>MANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vocabulary</td>
</tr>
<tr>
<td></td>
<td>F(2,83)</td>
</tr>
<tr>
<td>Novel Words</td>
<td>7.20**</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllogisms</td>
<td>8.06**</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary Details</td>
<td>5.58**</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary Main Ideas</td>
<td>4.98**</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary Themes</td>
<td>6.20**</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary Words Written</td>
<td>3.40*</td>
</tr>
</tbody>
</table>

NOTE: * p < .05, ** < .01, *** p < .001, † < .09
Academic Outcomes

Lastly, to investigate the academic outcomes of each group, I examined the representation of students enrolled in the two levels of studies for each group in Model 4. A chi-square test of independence indicated that the level of study in which students were enrolled was significantly associated with comprehension skill group, $\chi^2 (2, N = 86) = 18.95, p < .001$, Cramér’s $V = .47$. Whereas the majority of unexpected good and expected average comprehenders (95% and 76% respectively) were enrolled in the higher level of study (academic), 61% of unexpected poor comprehenders were enrolled in the lower level of study (applied).

Discussion

The purpose of this study was to investigate the characteristics of unexpected poor comprehenders, taking into account several limitations of previous research. Specifically, I first examined whether different control measures led to the identification of different individuals, and whether the groups differed on the variables that were not controlled. Second, I examined the characteristics of unexpected poor comprehenders for individual differences, word-related, and comprehension-related skills while controlling for word-reading accuracy, phonological decoding, reading rate, nonverbal ability, and vocabulary. Finally, I examined whether the results depended upon the nature of the contrast group used (expected average and unexpected good comprehenders).

Use of Different Selection/Control Criteria

I explored the effects of using five different control variables (word reading accuracy, phonological decoding, reading rate, nonverbal ability, and vocabulary) in different regression sets to identify unexpected poor comprehenders and two contrast groups (expected average and
unexpected good comprehenders). To my knowledge, this is the first study to compare the effects of using different selection criteria to identify unexpected poor comprehenders and contrast groups. Further, the inclusion of isolated word reading accuracy and reading rate to assess reading ability addresses the concern I raised about previous studies regarding the reading measure used to match groups (i.e., reading accuracy and reading comprehension were not always assessed independently, and an assessment of reading efficiency has not been regularly included).

I was successful in matching groups on the control measures that were explicitly entered for all but the model that included vocabulary (an issue we return to below). With the exception of phonological decoding, I found that if a variable was not explicitly controlled in the regression set, between-group differences on it persisted. Inclusion of the extra control variables is important because they eliminate possible interpretations of the sources of comprehension differences. The finding that my groups did not differ on phonological decoding, though it was not explicitly controlled, is consistent with previous findings that decoding no longer contributes unique variance once word reading skill has been taken into account (e.g., Goff et al., 2005).

The individuals identified in the three comprehension groups depended to some extent upon the selection/control criteria. I found substantial, but far from complete overlap in individuals identified as unexpectedly poor, expected average, or unexpected good comprehenders across models. Although very few individuals changed group membership across models (because of the buffer zones between groups), these differences were sufficient that significant between-group differences were evident on the control variables not entered into the regression models.
The regression technique to identify groups offers a methodological solution to concerns I raised in the literature review regarding the identification of unexpected poor comprehenders in previous studies. The regression technique allowed multiple dimensions to be taken into account (several reading and cognitive abilities) and in future studies could be used to control a wide range of characteristics. It also allowed for relative discrepancy between word reading and reading comprehension to be considered, so that a full range of students could be included. It also facilitated the differentiation of two contrast groups that have been combined in previous studies.

The finding that between-group differences persisted on the variables not entered into the regression set raises questions about many previous findings regarding unexpected poor comprehenders. Specifically, some of the differences that have been observed in previous studies between unexpected poor and more skilled comprehenders may have been due to factors that were controlled in this study but not in previous ones. For example, findings of weaknesses in working memory capacity in unexpected poor comprehenders (see Carretti et al., 2009, for a review) may be linked to the variance it shares with both word reading rate (Walczyk, Wei, Griffith-Ross, Goubert, & Cooper, 2007) and nonverbal (fluid) intelligence (Martínez & Colom, 2009).

When comparing the unexpected poor comprehenders to expected average and unexpected good comprehenders, I found that the majority of the effects observed were due to the superior skills of unexpected good rather than expected average comprehenders. This suggests that many of the differences reported in previous studies may have been unduly affected by the inclusion of unexpected good comprehenders in the comparison group (e.g., Cain, 2006;
Cain et al., 2001; Nation & Snowling, 1998a). Although the unexpected good comprehenders are an interesting group in themselves, the expected average comprehenders are a more conservative comparison group for evaluating what is different about the unexpected poor comprehenders.

**Characteristics of Different Comprehension Groups**

There is no shortage of theoretical frameworks within which the difficulties faced by unexpected poor comprehenders can be considered. Two current theories, the lexical quality hypothesis (Perfetti, 2007) and the construction-integration model (Kintsch, 1998) led me to expect that comprehension difficulties could occur due to weaknesses in the quality of word level knowledge, or at any of several levels of comprehension. Previous research has attributed the comprehension difficulties of children with age-appropriate word reading skill to weaknesses in word meaning (vocabulary) skills (e.g., Nation & Snowling, 1998b; Ricketts et al., 2007), and to higher-level weaknesses in inference making (see Cain & Oakhill, 2007, for a review).

Interpretation of the source difficulties of unexpected poor comprehenders has been made difficult owing to the limitations outlined in my review of the literature. Specifically, previous findings have been based on the use of different selection/control procedures, have not made a distinction in the nature of the comparison group (i.e., expected average and unexpected good comprehenders), and, to my knowledge, have not included the range of control measures, word-related, individual differences, and reading comprehension measures that were included in this study.

In my analyses that included all five controls, no group effects were found for the individual differences or word-related measures, indicating that the reading comprehension difficulties of unexpected poor comprehenders cannot be attributed to these variables. However,
it is important to note that there had been more differences before vocabulary was covaried. Vocabulary knowledge is pivotal for reading comprehension, whether it is seen as an area of weakness in unexpected poor comprehenders or instead as a skill on which they must be matched before comparisons are made.

For the comprehension-related variables, results depended upon the specific task, the comprehension level assessed, the comparison group used, and whether or not vocabulary was controlled (see Table 3.7). Without vocabulary controlled, all variables showed between-group differences, but many of these differences were between the unexpected poor and unexpected good comprehenders, or between the expected average and unexpected good comprehenders. When vocabulary was controlled, most of these differences disappeared, leaving mainly differences between unexpected poor and expected average comprehenders, which are the differences that I argue show most clearly the weaknesses in the unexpected poor comprehenders.

With vocabulary controlled, the unexpected poor comprehenders performed worse than the expected average comprehenders at the detail and main idea levels of the summary writing task. These differences, combined with the lack of differences on other variables tapping similar levels of comprehension processes (for example, the novel words task assesses detail processing, and the syllogisms task requires main idea and situation model construction), argue for the source of the difference being related to the summary task itself. Although it is admittedly post hoc and speculative, let us consider the nature of the summary task to suggest possible sources of the unexpected poor comprehenders’ difficulties. This task involved a longer text (680 words) than the other tasks, and it required learning more explicitly. Although other tasks involved
individual components involved in summary writing (e.g., working memory was assessed in the working memory measure, situation model construction in syllogisms), the summary task was more complex because it involved all of these components. One possibility is that the complexity of the task meant that weaknesses in individual components could not be compensated for with other components, and could contribute to weaknesses in the functioning of other components. For example, slight difficulties in understanding the meaning of a few words could impede microproposition formation, leading in turn to problems with main ideas. Weaker main idea comprehension could in turn impede re-construction of related details.

A second and related interpretation is that the underlying problem involves weaknesses in long-term memory and slower or less efficient processes in forming long-term memories, and that these weaknesses may be exacerbated when memory load increases (i.e., the longer text and the text-absent summary writing procedure). Supporting evidence for weaknesses in long-term memory was provided by Oakhill (1984), who found that unexpected poor comprehenders performed worse on a literal comprehension task when administered in a text absent format, but not in a text present format. Oakhill suggested that the effect could be attributed to their failure to produce appropriate strategies in organizing and retrieving information, but it could also be caused by more fundamental problems in forming associations. If unexpected poor comprehenders have difficulties at the detail level (i.e., in literal comprehension), this could affect the identification and construction of main ideas. The only measures used in this study that explicitly addressed strategies were the deep and surface approaches to learning; presumably the deep approach, and not the surface approach, would be required for the construction of main ideas and themes. Although there was a trend for the unexpected poor comprehenders to report
more surface and less deep processing, these effects were not significant, especially when vocabulary was controlled. Self-reported strategies may not be sufficient to capture the effect, and habitual strategies may change under complex task demands. Future research should investigate the long-term memory processes and strategies of unexpected poor comprehenders further.

**Effect of comparison groups**

Whereas previous studies have compared unexpected poor comprehenders to an undifferentiated group of more successful comprehenders, in cases with higher than normal reading comprehension ability (e.g., Cain, 2006; Cain et al., 2001; Nation & Snowling, 1998a), I compared them to average and higher-performing groups. If the purpose is to identify the particular weaknesses of the unexpected poor comprehenders, then average comprehenders are the more conservative comparison group. Whereas I found many variables on which the unexpected poor and unexpected good comprehenders differed in the MANOVAs and follow-up tests, there were relatively few differences between unexpected poor and expected average comprehenders. Controlling vocabulary in the MANCOVAs eliminated many of the differences between unexpected poor and unexpected good comprehenders, suggesting that a critical advantage of the unexpected good comprehenders lies in vocabulary. For diagnosing the nature of the weaknesses of unexpected poor comprehenders, the key results are that they performed worse than the expected average comprehenders on the two lower levels (details and main ideas) of the summary task.

**Unexpected good comprehenders**
In addition to the unexpected poor comprehenders who have been the focus of much previous research, we also identified a group of unexpected good comprehenders, those who performed above the level predicted from their word reading and nonverbal abilities. As with the unexpected poor comprehenders, the most useful comparison for uncovering the distinctive character of this group is with the expected average comprehenders. Before vocabulary was controlled, there were differences favoring the unexpected good comprehenders in print exposure, morphological awareness, novel words, syllogisms, and summary themes. All but the difference in themes disappeared once vocabulary was controlled. This suggests that one of the main advantages of these students is their vocabulary knowledge, consistent with the lexical quality hypothesis (Perfetti, 2007). However the theme difference is intriguing because this implicates the construction of situation models, learning, and deeper processing (Kintsch, 1998; Kirby & Woodhouse, 1994). These results suggest that the unexpected good comprehenders retain as many details and main ideas as the expected average comprehenders, but are more able or willing to go beyond that information to develop a more abstract or reflective perspective on the text. Future research should examine this group of students in greater detail, investigating the means to encourage this level of performance in other students.

Limitations and Future Directions

There are certain limitations to the present study that need to be acknowledged and addressed in future studies. First, although my results suggest that the reading comprehension difficulties of unexpected poor comprehenders cannot be attributed to word level knowledge, nonverbal ability, or my individual differences variables, I recognize that my measures may not have been sensitive enough to detect such differences. Furthermore, it is not clear whether
vocabulary should be a controlled variable, so that unexpected poor comprehenders would have to be average in vocabulary, or an outcome variable: my results indicated that vocabulary was powerful either way. Second, my measure of vocabulary was also a reading measure, one that only required the recognition of suitable synonyms for isolated words. I was, therefore, unable to determine how my results may have varied had an oral measure of vocabulary been used (e.g., Ouellette, 2006), or one which assessed the quality (depth) of word knowledge (Verhoeven & Perfetti, 2011). Further study is required to more fully understand the role of lexical size and the quality of semantic representation in relation to the comprehension difficulties of unexpected poor comprehenders. Third, although my results suggest that unexpected good comprehenders have strength in building an appropriate situation model, I did not test transfer of learning, or longer-term retention. I would expect students with more elaborate summaries at the main idea and theme levels to have better situation models of the text, and thus more knowledge that could be used subsequently and productively. For example, transfer could be demonstrated in terms of knowledge about the topic (e.g., making decisions about pest control in other situations), about structurally similar arguments (e.g., other texts about solving natural problems), or about scientific/argumentative texts in general. Fourth, unexpected poor comprehenders’ weaknesses at textbase and situation model levels were assessed only in the summary writing task; these weaknesses need to be explored with more varied tasks.

Conclusions

These results point to three main findings regarding the level at which reading comprehension difficulties occur for unexpected poor comprehenders. First, at the level of word-knowledge, a critical disadvantage of unexpected poor comprehenders lies in their weakness in vocabulary. Second, comprehension difficulties remain when vocabulary is controlled. Third, the
nature of these comprehension difficulties depends upon the comparison group used.

Comprehension difficulties at both the textbase and situation model levels remain when the unexpected poor comprehenders are compared to unexpected good comprehenders, whereas comprehension difficulties are limited to textbase level comprehension processes when they are compared to expected average comprehenders.

This study was carried out with high school students. Further insights may be provided by longitudinal studies beginning with a younger population, to investigate how the characteristics we observed originate. I also see a need for studies involving transfer of learning as opposed to simply comprehension, to observe the consequences of the characteristics we have observed for later learning and school performance.
References


Purpose
Although Study 2 identified a group of unexpected poor comprehenders when they were in Grade 10, two crucial questions were not examined. The first concerns whether their poor reading achievement was apparent at an earlier point in time (i.e., at Grades 3 or 6). The second question concerns how well they fared on the government mandated test (OSSLT) administered in Grade 10. These two questions are addressed in the analysis that follows, and serve to provide a bridge between Study One and Two of my dissertation.

Method
Subjects who were identified in Model 4 of Study Two as unexpected poor, expected average, and unexpected good comprehenders, and for whom achievement data on government mandated reading tests in Grades 3, 6, and 10 were available, were used in the analysis. Performance for each group for these testing periods was examined in the following manner. For Grades 3 and 6, performance for individuals in each group was based on their scores at one of the four provincial achievement levels, with level 3 indicating that provincial expectation was met. For the Grade 10 OSSLT, performance for individuals in each group was based on a pass or fail criterion.

Results

Of the unexpected poor comprehenders identified in Model 4 of Study Two, data were available in Grades 3, 6, and 10 for 32%, 60%, and 89% of the group (N=28); expected average comprehenders (N=37), 62%, 84%, 97%, respectively; unexpected good comprehenders (N=21), 71%, 76%, 95%, respectively.
The results are outlined in Table 4.1 and indicate that a higher percentage of unexpected poor comprehenders could be identified as achieving at the lowest level of attainment (Level 1 Grades 3 and 6, and failing OSSLT) than expected average or unexpected good comprehenders. The percentage of unexpected poor comprehenders who did not meet provincial expectation, compared to unexpected good comprehenders, was consistently higher: levels 1 and 2 Grades 3 and 6, fail OSSLT (66.6, 52.9, 24 versus 20, 25, 5, respectively). For Grade 3, a greater percentage of unexpected poor than expected average comprehenders did not meet the provincial standard of level 3 (total percentage at levels 1 and 2, 66.6 versus 52.2, respectively); for Grade 6 the percentage was 52.9 versus 54.9, respectively and for the OSSLT, 24% versus 8.3% respectively were not successful. None of these differences, however, were found to be significant (Grade 3, \( \chi^2(1, N = 32) = 0.552, p = .46 \); Grade 6, \( \chi^2(1, N = 48) = 0.016, p = .90 \); OSSLT, \( \chi^2(1, N = 61) = 2.878, p = .09 \))

**Discussion and Conclusions**

Although there is a tendency for the unexpected poor comprehenders to have performed worse on the government-mandated assessments in Grades 3, 6, and 10, it is also clear that a substantial number of them did pass those assessments. This suggests that some of them have long-standing difficulties, perhaps in the areas of language or vocabulary that would allow them to be identified much earlier. However, others seem to be more consistent with the notion of having a “late-emerging” difficulty, having shown good performance at the Grade 3 and 6 assessments. In that no significant differences were noted between the unexpected poor and expected average comprehenders in Grade 3 and 6, supports the idea of late emergence, but the non-significant results of the Grade10 EQAO test says further that it isn't always seen in
curriculum-based measures. A possible explanation may rest in the inclusion of the writing component in the OSSLT results. Further research is required in order to relate unexpected poor comprehenders’ status to real performance in school.

It should be noted that that a considerable number of unexpected poor comprehenders passed the OSSLT, which was given in the same year as the CTBS reading comprehension test on the basis of which the comprehension groups were formed. To some extent this may show non-equivalence of the two measures, but it may also be due to the way in which the regression method identified unexpected poor comprehenders (i.e., relative to their performance on predictor variables, not in terms of an absolute criterion score).

Three conclusions can be drawn from this analysis. First, that almost half of the unexpected poor comprehenders could be identified as poor comprehenders as early as Grade 3 suggests that their reading problems are not entirely “late emerging” (Leach, Scarborough, & Rescorla, 2003). Second, that almost one-quarter of the unexpected poor comprehenders were unsuccessful on the OSSLT indicates that their poor comprehension has negative consequences for learning outcomes in school settings. Third, a substantial number of unexpected poor comprehenders in Grade 10 had performed well on the earlier assessments in Grades 3 and 6; this suggests that efforts continue to identify such students and to improve their reading comprehension skills.

References
Table 4.1

*Achievement of Unexpected Poor (UPC), Expected Average (EAC), and Unexpected Good Comprehenders (UGC) in Grades 3, 6 and 10*

<table>
<thead>
<tr>
<th>Grade</th>
<th>Level</th>
<th>n</th>
<th>percent</th>
<th>n</th>
<th>percent</th>
<th>n</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 3</td>
<td>1</td>
<td>2</td>
<td>22.2</td>
<td>2</td>
<td>8.7</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>44.4</td>
<td>10</td>
<td>43.5</td>
<td>3</td>
<td>20.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>22.2</td>
<td>10</td>
<td>43.5</td>
<td>8</td>
<td>53.3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>11.1</td>
<td>1</td>
<td>4.3</td>
<td>4</td>
<td>26.7</td>
</tr>
<tr>
<td>Grade 6</td>
<td>1</td>
<td>5</td>
<td>29.4</td>
<td>2</td>
<td>6.5</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>23.5</td>
<td>15</td>
<td>48.4</td>
<td>4</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7</td>
<td>41.2</td>
<td>14</td>
<td>45.2</td>
<td>10</td>
<td>62.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>5.9</td>
<td>0</td>
<td>0.0</td>
<td>2</td>
<td>12.5</td>
</tr>
</tbody>
</table>

**OSSLT**

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>19</td>
<td>76.0</td>
<td>33</td>
<td>91.7</td>
<td>19</td>
<td>95.0</td>
<td></td>
</tr>
<tr>
<td>Fail</td>
<td>6</td>
<td>24.0</td>
<td>3</td>
<td>8.3</td>
<td>1</td>
<td>5.0</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 5: DISCUSSION

Introduction

There is evidence that a substantial number of high school students have not achieved acceptable levels of reading comprehension (OECD, 2010). It has also been argued that little research has been devoted to understanding the nature of this age group’s reading difficulties (e.g., Carnegie Council on Advancing Adolescent Literacy, 2010). Therefore, the overall purpose of this dissertation was to increase our understanding of the factors that underlie the poor reading comprehension abilities in this group. It did so by reporting the results of two studies that utilized current models and theories of reading comprehension in a manner that would inform educational research and practice.

The main aim of Study One was to examine the nature of growth patterns in the development of reading achievement between Grades 3 and 10. Three models describing the development of reading achievement over time were used: stability (Cunningham & Stanovich, 1997), cumulative growth (Leppänen, Niemi, Aunola, & Nurmi, 2004; Stanovich, 1986), and compensation (Leppänen et al., 2004; Parrila, Aunola, Leskinen, Nurmi, & Kirby, 2005; Phillips, Norris, Osmond, & Maynard, 2002). Based on their reading performance, children were classified into four achievement groupings (i.e., poor, below-average, average, and above-average). To address limitations of previous research, tests investigating statistical effects associated with regression toward the mean were conducted. The results indicated a considerable degree of stability in reading comprehension across the period from grade 3 to grade 10.

The main aim of Study Two was to examine the nature of comprehension difficulties in high school aged unexpected poor comprehenders. In addressing the methodological limitations of previous research with unexpected poor comprehenders, the theoretical basis of Perfetti’s
lexical quality hypothesis (2007) and Kintsch’s construction-integration model (1998) was used to examine at what level (i.e., word knowledge, textbase, situation model) reading comprehension difficulties occurred in this group. The study explored the implications of including a wide range of control variables and of making a distinction between two comparison groups, expected average and unexpected good comprehenders. The individuals identified in the three comprehension groups depended to some extent upon the selection/control criteria. Although there was substantial overlap in individuals identified as unexpectedly poor, expected average, or unexpected good comprehenders across models, these differences were sufficient that significant between-group differences were evident on the control variables not entered into the regression models. Vocabulary was associated with poor reading comprehension, but with it controlled statistically, the main deficiencies of the unexpected poor comprehenders were in identifying details and main ideas in summary writing; thus the unexpected poor comprehenders showed weaknesses at the word (vocabulary) and textbase or situation model (details and main ideas) levels. The unexpected good comprehenders, on the other hand, showed an advantage in vocabulary, and with that controlled, the only difference between them and the expected average comprehenders was at the theme level of the summary writing task, a measure of situation model construction or deeper processing.

The remainder of this discussion is presented in four sections. The first three sections elaborate on three broad themes that arise from my dissertation findings: the nature of the reading comprehension construct, the assessment of reading comprehension, and the identification and remediation of those who struggle with reading comprehension. The
implications for educational practice and future research are incorporated into each thematic section. The conclusion follows in the final section.

The Reading Comprehension Construct

The first theme that arises from my dissertation concerns the nature of the reading comprehension construct. That reading comprehension is complex has long been recognized. Indeed, over 100 years ago Edmund Huey wrote, “to completely analyze what we do when we read, would almost be the acme of the psychologist’s achievements, for it would be to describe many of the most intricate workings of the human mind.” (1968/1908). Current definitions acknowledge that effective reading comprehension involves the intentional, active construction of meaning that is appropriate to the type of text and purpose of reading (e.g., National Reading Panel, 2000; Snow, Burns, & Griffin, 1998; Torgesen, Houston, Rissman, Decker, Roberts, Vaughn, et al., 2007). A wide variety of skills and knowledge support effective reading comprehension. Torgesen and colleagues (2007) identified a number of areas of student knowledge and skill that must continue to grow from 4th to 12th grades in order for students to maintain at least grade level reading skills across that development span: sight reading vocabulary must be extended to unfamiliar words in increasingly difficult text (fluency), vocabulary or knowledge of words (depth and breadth) must expand dramatically, conceptual knowledge and understanding must grow, higher level thinking and reasoning and self-regulated use of reading comprehension strategies must develop, and motivation and engagement in broad and deep reading must be maintained or acquired for understanding and learning from text. These skills and strategies vary with age, reading experience, quality of instruction, context and motivation. As a result, both the processes and the products of reading comprehension are
multidimensional, developmental, and variable. Thus, it is not surprising that “reading comprehension is difficult to define simply and to measure neatly” (Paris, n.d., p. 1).

As emphasized in the introductory chapter of my dissertation, theories are required to make sense of this complexity, and current theories emphasize the multi-dimensional nature of the processes and products underlying effective reading comprehension (e.g., Graesser & McNamara, 2011; Kintsch & Kintsch, 2005; Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007). Indeed, that reading comprehension is considered to be multi-dimensional suggests that it may be more appropriate to refer to the reading comprehension construct in the plural form (constructs), rather than in singular form.

In recognition of the complexity and multi-dimensional nature of the reading comprehension construct, Study Two of my dissertation used a theoretically guided multi-component approach that was theoretically guided to investigate the sources of reading comprehension difficulties in a group of poor readers. The lexical quality hypothesis (Perfetti, 2007) and the construction-integration model (Kintsch, 1998) were utilized as a basis for examining at what level comprehension difficulties can occur (word knowledge, textbase, situation model level). These levels can be regarded as parts of, but not the entire spectrum, of the dimensions that comprise the reading comprehension construct and in which comprehension difficulties may occur.

My findings provide support for the multi-dimensionality of the reading comprehension construct in the following ways. A group of poor comprehenders was identified whose reading achievement could not be attributed to poor word reading skills (i.e., decoding, word reading accuracy, and reading rate) or general ability. Rather, one of their weaknesses occurred at the
semantic level of word knowledge (i.e., vocabulary). This finding is consistent with the tenets laid out in the lexical quality hypothesis – that is, variation in the quality and unity of the form (phonology, orthography) and meaning (vocabulary) of word representations has consequences for reading skill, including comprehension (Perfetti, 2007). Further support for the multi-dimensionality of reading comprehension was found in that weaknesses exhibited by this group extended beyond lower word knowledge skills to comprehension difficulties at the textbase and situation model levels outlined in Kintsch’s (1998) construction-integration model. These findings are consistent with the notion that word level skill and knowledge is necessary, but insufficient for effective reading comprehension. Importantly, the findings in Study Two that reading comprehension difficulties can be attributed to weaknesses at different levels of comprehension (i.e., vocabulary, textbase, situation model) indicates that reading comprehension difficulties of adolescent students cannot be reduced to a single source. From an educational perspective, this has critical implications for the manner in which reading comprehension achievement should be assessed. The assessment of reading comprehension is the second theme that arises from my dissertation studies, and to which I now turn.

The Assessment of Reading Comprehension

The assessment of reading comprehension has been controversial (Pearson & Hamm, 2005). At the center of this controversy rests incongruence between the insights provided by theory and empirical evidence concerning the nature of this construct and the manner in which it commonly is assessed. Theory and empirical evidence have demonstrated that reading comprehension is variable and multi-dimensional. Yet, the traditional reliance has been on tests with little variation in the material read and relatively little variation in the response formats used within tests (Sweet, 2005). This is in contrast to “cognitive approaches [that] acknowledge that
measures of reading comprehension are variable and indirect indicators of reading comprehension…that serve as formative measures that are useful for instruction and remediation” (Paris, n.d.). Further, because commonly used measures of reading comprehension report achievement in uni-dimensional fashion (one score), they “inadequately represent the complexity of the target domain …. conflating comprehension with vocabulary, domain-specific knowledge, word reading ability, and other reader capacities involved in comprehension”(RAND Reading Study Group, 2002).

My Study One used government-mandated or traditional measures that provided a single score indicating reading comprehension ability. The single scores were the basis of the groupings that were shown to be stable over the grade 3 to 10 period. My Study Two followed the cognitive approach, not only employing theory-based measures of comprehension components or levels, but also relying upon other measures of various cognitive abilities or processes. In contrast to the first study, the second one emphasized the complexity of the reading comprehension construct and the skills on which it depends; little of this could have been accomplished with only single-score omnibus measures of reading comprehension.

To be fair, commonly used tests serve the purpose for which they were intended: the quantitative scores on the same scales provide summative measures of reading that can be used to sort and compare students (e.g., Cross & Paris, 1987; Sweet, 2005). Both the EQAO and Canadian Test of Basic Skills used in Study 1 and 2 are examples of tests that serve this purpose. Reporting reading achievement as single scores may allow one to identify which students struggle with reading comprehension as was done in my Study 1. However, given the multi-dimensional nature of reading comprehension, they do not provide information regarding the
source of reading difficulties. An alternate purpose for assessment is that it serves a diagnostic function. This is needed to ensure that remediation efforts are appropriately directed. As such, the inclusion of the control variables in Study 2, indicated that word reading was not a source to target for remediation for this group of poor comprehenders.

The RAND Reading Study Group (2002) suggested that diagnostic assessment could initially serve to identify subtypes of poor comprehenders in terms of the components (e.g., decoding and listening comprehension) and the desired outcomes of comprehension (e.g., engagement in reading, and learning). However, the ultimate purpose of diagnostic assessment is not to simply identify subtypes of poor comprehenders (e.g., students with learning disabilities, unexpected poor comprehenders). Rather, the information provided should serve to inform educational practitioners of the specific domains that instruction should target. Results from my Study 2 indicate that both vocabulary and higher levels of comprehension are areas of weakness in unexpected poor comprehenders that could be targeted for instruction. Ideally, diagnostic assessment should aim to provide teachers with a profile of reading skills and deficits of individual students, thereby requiring that a broad range of reading skills, knowledge and attitudes be assessed.

For younger pre-school and early elementary school children, a second purpose of assessment is to screen children who, without intervention, would be likely to develop later reading problems. Such assessments, administered in oral and visual form, should incorporate measures that are strongly predictive of the reading or cognitive components that contribute to the effective development of reading skill (e.g., phonemic awareness and letter knowledge to word reading skill). But reading skill, as emphasized in the simple view of reading, is only part
of the story in understanding the development of skill in reading comprehension. It is, therefore, unfortunate that research on reading development in this age group has focused largely on the role of basic literacy skills (i.e., phonological awareness, letter knowledge, and vocabulary). In contrast, little research has been carried out on the development of comprehension skill at these ages (for research that has been done see for example, van den Broek, 1997; van den Broek, Kendeou, Kremer, Lynch, Butler, White et al., 2005; Cain & Oakhill, 1999; Paris & Paris, 2003). As a result, in comparison to the availability of screening and diagnostic assessments in basic literacy skills, fewer assessments of comprehension skills are available for this age group. Given that comprehension skills and basic literacy skills develop in tandem (simultaneously and independently), rather than sequentially (Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007), it is important that more research be devoted to the development of measures that can provide early identification of children (i.e., unexpected poor comprehenders) whose weaknesses extend beyond the word reading level to higher-order comprehension weaknesses (Kendeou, Lynch, van den Broek, Espin, White, & Kremer, 2005).

Beyond diagnosis and screening, a third purpose of assessment is to monitor student progress (e.g., responsiveness to intervention) in terms of the intended learning outcomes. Research has shown that students show significant gains in response to theoretically motivated interventions that are targeted at the source of reading comprehension deficits (e.g., decoding, vocabulary, text comprehension) (e.g., Clarke, Snowling, Truelove, & Hulme, 2010; Scammacca, Roberts, Vaughn, Edmonds, Wexler, Reutebuch, et al., 2007). However, these gains have been more commonly observed and are greater when research developed tests have been used to assess outcomes, rather than commercially available norm-referenced tests. Further,
transfer effects of intervention have been observed on some, but not all, standardized reading comprehension measures. This reinforces the notion that different tests are not comparable in the skills they assess (e.g., Keenan, Betjemann, & Olson, 2008). Importantly, standardized tests may not reflect the learning outcomes of reading comprehension that are critical for school and workplace success (i.e., knowledge acquisition, engagement in reading). Beyond the issues associated with the choice of measure used to monitor student progress on intended learning outcomes, “when response to a well-founded intervention is poor, it is important to consider the possibility that co-occurring difficulties are affecting progress and may require separate treatments” (Snowling & Hulme, 2011, p.19). As such, a new cycle of diagnostic assessment may be necessary.

In summary, it is clear that the reliance on one assessment measure is not capable of achieving these varied purposes. The RAND Reading Study Group (2002) argued that a system of assessment measures is required, and identified a number of requirements of an ideal system that my discussion has not yet addressed. These include but are not limited to the following: sensitivity to intra-individual differences (e.g., differences in performance across activities with varying purposes and with a variety of texts and text types, capacity to reflect authentic outcomes (e.g., students’ beliefs about reading and their performance as readers); developmental sensitivity (e.g., to the relationship of not only word reading, but also oral language comprehension in the development of reading comprehension). My Study Two shows the value of employing theory-based models of the reading comprehension process (e.g., those of Kintsch, 1998, and Perfetti, 2007) and examining student performance within their context. It is clear that an aggressive research agenda is required to meet the needs of developing more adequate assessments of
reading comprehension. Until that time, much can still be garnered from research that seeks to inform educational practice of reading comprehension difficulties with the assessment measures that presently exist. This brings me to the third and final theme of my dissertation.

**Identification and Remediation of Reading Comprehension Difficulties**

The third theme concerns the identification and remediation of reading difficulties. The overall pattern of stability exhibited in students’ reading achievement in Study One indicates that children identified in the lowest achievement groupings (poor and below average) in the early grades do not outgrow their reading problems. Further evidence of long-standing difficulties was found in the analyses provided in Chapter 4 that indicated a higher proportion of unexpected poor comprehenders had been achieving at the lowest levels of attainment (Level 1 in Grades 3 and 6, failing the OSSLT) than expected average or unexpected good comprehenders. From an educational perspective, this indicates that adolescent students who struggle with effective reading comprehension in later years could be identified early in their school careers.

As outlined in Chapter 1, initial skill in reading comprehension draws on oral language factors that contribute to both efficient word reading and comprehension skill. Initial weaknesses in these domains set the stage for further difficulties at later ages (Snowling & Hulme, 2011). Early identification, with assessments that are sensitive to the language factors that contribute to the development of not only word reading, but also to comprehension skill, is, therefore, important so that effective remediation can be provided at the earliest time possible. In relation to my dissertation, Study Two investigated a group of poor comprehenders who did not have word-reading problems, and whose reading comprehension difficulties became more apparent later in the elementary school years. These unexpected poor comprehenders had low vocabulary
knowledge by grade 10. Lacking earlier measures of vocabulary, it is not possible to know when their vocabulary skills began to lag behind those of their peers: they may have had low vocabulary skills from the time of school entry, pointing to a need for more stimulating early environments. Their low vocabulary scores may also have been due to a lack of reading experience, which may in turn have been due to poor reading comprehension skills. These unexpected poor comprehenders also showed weaknesses on the detail and main idea levels of the summary writing task, raising hypotheses about difficulties in selecting important information in texts or in constructing situation models. These findings point to a need for research to investigate the degree to which these weaknesses can be remediated, and if they can, whether there are consequential improvements in reading comprehension.

Addressing issues involving the timing, type, and intensity of intervention is crucial to closing the gap in reading achievement. Remediation that is delayed until children are older may be more difficult (Roberts, Torgesen, Boardman, & Scammacca, 2008; Shaywitz, Morris, & Shaywitz, 2008). Supporting this view, a meta-analysis of 31 intervention studies, grouped into five intervention types (i.e., fluency, vocabulary, reading comprehension strategies, word study, or multiple components of reading instruction) was conducted with Grades 4-12 students identified as struggling readers (Scammacca et al., 2007); the vocabulary and word study interventions are relevant to the findings of Study Two about vocabulary weaknesses. Intervention effects, for all five intervention types, were significant for reading comprehension outcomes (researcher developed and standardized measures) in studies when participants were middle grade students (Grades 4-8) ($d = 1.11$) as opposed to high school students ($d = .59$). The pattern repeated itself when only standardized measures were used, but the magnitude of effect
sizes was lower ($d = .47$ and .14). The finding that intervention is more effective when provided as early as possible is consistent with research findings at the primary grade level (for a review, see Snowling & Hulme, 2011). Therefore, not providing effective instruction and remediation early means that as these students progress in their schooling, the more ground they will have to cover, that, in turn, will affect the intensity and duration of the required intervention.

There is also a need to specify the kinds of intervention programs that will be most effective (Snowling & Hulme, 2011) for struggling adolescent readers (Biancarosa & Snow, 2004; Deshler, 2006). Three of the instructional topic areas selected by The National Reading Panel (2002) for intensive study, as critical in developing effecting reading comprehension skill, are relevant here: alphabetics (i.e., phonemic awareness and phonics instruction), fluency, and comprehension (which included vocabulary and text comprehension instruction). Measures for each of these areas were included in Study 2. Systematically using the control variables (word reading, reading rate, and vocabulary) in this study allowed me to determine which areas may benefit from, and hence be targeted (vocabulary and comprehension), for remediation.

The National Reading Panel (2002) found evidence for the effectiveness of seven areas of comprehension instruction, including comprehension monitoring (readers learn how to be aware of their understanding of the material), cooperative learning (students learn reading strategies together), use of graphic and semantic organizers (readers make graphic representations of the material to assist comprehension), question answering (readers answer questions posed by the teacher and receive immediate feedback), question generation (readers ask themselves questions about various aspects of the story), story structure (students are taught to use the structure of the story to help them recall story content to answer questions about what they have read), and
summarization (readers are taught to integrate ideas and generalize from the text information). It should be noted that most of these are relevant to the weaknesses identified in Study Two regarding details and main ideas in the summary writing task.

A synthesis of 13 reading interventions and their effects on reading comprehension outcomes by Edmonds and colleagues (2009) with Grades 6-12 struggling readers reported that improvements to reading comprehension outcomes were greatest when intervention targeted comprehension skills and strategies (such as those identified by the National Reading Panel, 2000), and multiple reading components (more than one component of reading, such as word study with comprehension or fluency with comprehension) \((d = 1.23 \text{ and } .72, \text{ respectively})\). The average effect size for word study \((d = .34)\) on reading comprehension outcomes was small. However, this finding is consistent with the notion that improvements in word reading may be necessary for those who still struggle at the word-level, but not sufficient for effective reading comprehension. By the time students reach high school, other factors, such as background knowledge, vocabulary, and use of active strategies and motivation to derive meaning at deeper levels (Kintsch & Kintsch, 2005) contribute to effective comprehension.

Of interest, was that null effects on reading comprehension outcomes were reported for reading fluency interventions in both the Scammacca meta-analyses and Edmonds synthesis \((d = -.07 \text{ and } -.03, \text{ respectively})\). Repeated reading, a technique that has been found to be effective in improving reading comprehension outcomes in younger students (National Reading Panel, 2000), was the most prevalent fluency intervention in the studies used in these studies. Scammacca and colleagues suggested that research using different intervention techniques is needed to determine how to effectively remediate fluency in adolescent struggling readers.
Others have suggested that the correlation between oral reading fluency and comprehension may change with development, decreasing steadily with age and text difficulty (Francis, Fletcher, Catts, & Tomblin, 2004; Paris, Carpenter, Paris, & Hamilton, 2004). It may also be that a portion of struggling readers included in these studies were unexpected poor comprehenders, a group illustrated in Study 2 whose weaknesses does not lie at the word level. Although limited by the low number of studies that included high school students, the primary findings from both of these synthesis are that older struggling readers can improve in their reading comprehension when taught reading comprehension strategies and practices.

Given that the source of reading difficulties is not the same for all struggling readers (Snowling & Hulme, 2011), there is a clear need to specify more clearly for whom which interventions work (Lyon & Moats, 1997). For high school students who lack word reading skills (who would be, in the sense of Study Two, expected poor comprehenders), it is necessary to build word-level skills while teaching comprehension so that access to increasingly difficult levels of print is available to them (e.g., Lovett, Lacerenza, De Palma, M., & Frijters, in press). Similarly, theoretically motivated interventions when targeted at those with specific comprehension deficits (i.e., vocabulary and text comprehension), at least with younger children, have been shown to produce significant gains on some, but not all, standardized reading comprehension measures (Clarke et al., 2010). The weaknesses observed in my results with older unexpected poor comprehenders strongly suggest that remediation with older students must be directed at vocabulary as well as at deeper levels of comprehension. Future research with this group is required to determine the efficacy of an approach that targets these areas and monitors students’ progress on a wide array of reading comprehension outcomes.
Of particular importance to the remediation of reading difficulties is that teachers are adequately trained to identify and remediate sources of reading difficulties in their students. To accomplish this, sustained professional development for teachers in instructional practices that promote effective reading comprehension is necessary. This appears to be particularly pertinent for high school educators who do not feel they have the time to include explicit comprehension instruction in their already crowded curricula (Scanlon, Deshler, & Schumaker, 1996), and who often believe that literacy is not within their area of competency or responsibility (O’Brien, Stewart, & Moje, 1995). Further, there is a need by educators to appreciate that there is no single “quick fix” for complex reading problems (Allington & Walmsley, 1995; Alvermann, 2002). Rather, effective approaches to literacy instruction must of necessity be custom-tailored to the needs of individual students, be comprehensive and multi-faceted, and integrated within and across curricula (Sturtevant & Linek, 2003).

**Conclusion**

Numerous studies have investigated the factors underlying the reading achievement of elementary school students. Of the available studies, only a small portion has been carried out with high school students. By addressing the methodological limitations associated with past reading comprehension research (i.e., restrictions in the range of age of sample, measurement error, multi-dimensionality of the reading comprehension construct, confounds associated with the use of different control variables, nature of comparison group), the studies included in my dissertation provide two major findings. First, later reading comprehension difficulties are evident early in children’s schooling. This finding emphasizes the need for early identification and remediation of reading comprehension difficulties. Second, reading comprehension difficulties was found to occur at multiple levels, specifically in word-level vocabulary
knowledge and in textbase and situation model construction. Because the data in Study 2 were
correlational, issues regarding the direction of causality between vocabulary and text
comprehension were not addressed. Further research regarding the direction of causality is
required so that remediation efforts can be appropriately informed. Specifically, it may be that
improvements to vocabulary lead to improvements in text comprehension. However, it may well
be that having poor comprehension leads to having low vocabulary, in which case teaching
vocabulary will not show improvements in text comprehension. The second finding of reading
comprehension occurring at multiple levels assists in the conceptualization of the multi-
dimensionality of reading comprehension, and highlights the need for reading comprehension
assessments and instructional practices that reflect the multi-dimensional nature of the reading
comprehension construct. The ability to identify specific deficits of students who struggle with
reading comprehension is crucial so that remediation efforts can be targeted appropriately.
Although intervention in the early school years is preferable, research evidence informs us that
high school is not too late in closing the achievement gap in reading achievement (Edmonds et
al., 2009; Scammacca et al., 2007; Torgesen et al., 2007). A crucial question for future research
concerns the means by which educators at the secondary school level can be provided with the
appropriate kinds of professional development necessary to achieve this goal.
References


Torgesen, J. K., Houston, D., Rissman, L., Decker, S., Roberts, G., Vaughn, S., Wexler, J.,
Francis, D., Rivera, M., & Lesaux, N. (2007). Academic literacy instruction for adolescents:
A guidance document from the Center on Instruction. Portsmouth, NH: RMC Research
Corporation, Center on Instruction. Available online at www.centeroninstruction.org.
comprehension from childhood to adulthood. In P. van den Brock, P.J. Bauer, & T. Bourg
(Eds.), Developmental spans in event comprehension and representation (pp. 321-342).
Mahwah, NJ: Lawrence Erlbaum Associates.
APPENDIX A: GENERAL RESEARCH ETHICS APPROVAL

June 18, 2006

Bozena Kwiatkowska-White
PhD Student
Faculty of Education
Queen's University

GREB Ref # GEDUC-281-06
Title: “Understanding reading comprehension in adolescent students”

Dear Ms. Kwiatkowska-White:

The General Research Ethics Board (GREB) has given expedited approval to your proposal entitled “Understanding reading comprehension in adolescent students”. In accordance with the Tri-Council Guidelines (article D.1.6) and Senate Terms of Reference (article G), your project has been approved for one year and is contingent upon relevant school board approval. At the end of each year, GREB will ask if your project has been completed and, if not, what changes have occurred or will occur in the next year.

You are reminded of your obligation to advise the GREB, with a copy to your unit REB, of any adverse event(s) that occur during this approval period (details available on our webpage www.queensu.ca/vpr/greb/addforms.htm). An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that any adverse events must be reported to the GREB within 48 hours.

You are also reminded that all changes that might affect human participants must be approved by the GREB. Examples of required approvals are: changes in study procedures or implementations of new aspects into the study procedures that affect human subjects. These changes must be sent to Linda Frid at the Office of Research Services or lfrid@post.queensu.ca prior to implementation. Ms. Frid will seek the approval of the GREB reviewer(s) who originally assessed your application.

On behalf of the General Research Ethics Board, I wish you continued success in your research.

Yours sincerely

Lee Rodrigar, PhD
Associate Professor and
Member, General Research Ethics Board

C.C.: Dr. K. Smithrim & D. Klinger, Co-Chairs Unit REB
Dr. John Kirby, Faculty Supervisor
Heather Cross
## APPENDIX B: EXPOSURE TO PRINT QUESTIONNAIRE

<table>
<thead>
<tr>
<th></th>
<th>Author</th>
<th></th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Douglas Adams X</td>
<td>35</td>
<td>Robert Emery</td>
</tr>
<tr>
<td>2</td>
<td>Lauren Adamson</td>
<td>36</td>
<td>Frances Fincham</td>
</tr>
<tr>
<td>3</td>
<td>Eric Amsel</td>
<td>37</td>
<td>Timothy Findley X</td>
</tr>
<tr>
<td>4</td>
<td>V. C. Andrews X</td>
<td>38</td>
<td>Martin Ford</td>
</tr>
<tr>
<td>5</td>
<td>Isaac Asimov X</td>
<td>39</td>
<td>Ram Frost</td>
</tr>
<tr>
<td>6</td>
<td>Margaret Atwood X</td>
<td>40</td>
<td>Howard Gardner</td>
</tr>
<tr>
<td>7</td>
<td>Margaritia Azmitia</td>
<td>41</td>
<td>Stephen J. Gould X</td>
</tr>
<tr>
<td>8</td>
<td>Russell Banks X</td>
<td>42</td>
<td>Andrew Greeley X</td>
</tr>
<tr>
<td>9</td>
<td>Oscar Barbarin</td>
<td>43</td>
<td>Frank Gresham</td>
</tr>
<tr>
<td>10</td>
<td>Reuben Baron</td>
<td>44</td>
<td>John Grisham X</td>
</tr>
<tr>
<td>11</td>
<td>Gary Beauchamp</td>
<td>45</td>
<td>Steven Hawking X</td>
</tr>
<tr>
<td>12</td>
<td>Thomas Bever</td>
<td>46</td>
<td>Frank Herbert X</td>
</tr>
<tr>
<td>13</td>
<td>Brian Bigelow</td>
<td>47</td>
<td>S. E. Hinton X</td>
</tr>
<tr>
<td>14</td>
<td>Elliot Blass</td>
<td>48</td>
<td>Charles Hulme</td>
</tr>
<tr>
<td>15</td>
<td>Judy Blume X</td>
<td>49</td>
<td>Robert Inness</td>
</tr>
<tr>
<td>16</td>
<td>Dale Blyth</td>
<td>50</td>
<td>John Jakes X</td>
</tr>
<tr>
<td>17</td>
<td>Hilda Borko</td>
<td>51</td>
<td>Robert Jordon X</td>
</tr>
<tr>
<td>18</td>
<td>Ann Brasharest X</td>
<td>52</td>
<td>Frank Keil</td>
</tr>
<tr>
<td>19</td>
<td>Dan Brown X</td>
<td>53</td>
<td>Stephen King X</td>
</tr>
<tr>
<td>20</td>
<td>P.E. Bryant</td>
<td>54</td>
<td>Dean Koontz X</td>
</tr>
<tr>
<td>21</td>
<td>Meg Cabot X</td>
<td>55</td>
<td>Judith Krantz X</td>
</tr>
<tr>
<td>22</td>
<td>Agatha Christie X</td>
<td>56</td>
<td>Louis L’Amour X</td>
</tr>
<tr>
<td>23</td>
<td>Douglas Adamson</td>
<td>57</td>
<td>Robert Emery</td>
</tr>
<tr>
<td>24</td>
<td>Lauren Adamson</td>
<td>58</td>
<td>Frances Fincham</td>
</tr>
<tr>
<td>25</td>
<td>Eric Amsel</td>
<td>59</td>
<td>Timothy Findley X</td>
</tr>
<tr>
<td>26</td>
<td>V. C. Andrews X</td>
<td>60</td>
<td>Martin Ford</td>
</tr>
<tr>
<td>27</td>
<td>Isaac Asimov X</td>
<td>61</td>
<td>Ram Frost</td>
</tr>
<tr>
<td>28</td>
<td>Margaret Atwood X</td>
<td>62</td>
<td>Howard Gardner</td>
</tr>
<tr>
<td>29</td>
<td>Margaritia Azmitia</td>
<td>63</td>
<td>Stephen J. Gould X</td>
</tr>
<tr>
<td>30</td>
<td>Russell Banks X</td>
<td>64</td>
<td>Andrew Greeley X</td>
</tr>
<tr>
<td>31</td>
<td>Oscar Barbarin</td>
<td>65</td>
<td>Frank Gresham</td>
</tr>
<tr>
<td>32</td>
<td>Reuben Baron</td>
<td>66</td>
<td>John Grisham X</td>
</tr>
<tr>
<td>33</td>
<td>Gary Beauchamp</td>
<td>67</td>
<td>Steven Hawking X</td>
</tr>
<tr>
<td>34</td>
<td>Thomas Bever</td>
<td>68</td>
<td>Frank Herbert X</td>
</tr>
<tr>
<td>35</td>
<td>Brian Bigelow</td>
<td>69</td>
<td>S. E. Hinton X</td>
</tr>
<tr>
<td>36</td>
<td>Elliot Blass</td>
<td>70</td>
<td>Charles Hulme</td>
</tr>
<tr>
<td>37</td>
<td>Judy Blume X</td>
<td>71</td>
<td>Robert Inness</td>
</tr>
<tr>
<td>38</td>
<td>Dale Blyth</td>
<td>72</td>
<td>John Jakes X</td>
</tr>
<tr>
<td>39</td>
<td>Hilda Borko</td>
<td>73</td>
<td>Robert Jordon X</td>
</tr>
<tr>
<td>40</td>
<td>Ann Brasharest X</td>
<td>74</td>
<td>Frank Keil</td>
</tr>
<tr>
<td>41</td>
<td>Dan Brown X</td>
<td>75</td>
<td>Stephen King X</td>
</tr>
<tr>
<td>42</td>
<td>P.E. Bryant</td>
<td>76</td>
<td>Dean Koontz X</td>
</tr>
<tr>
<td>43</td>
<td>Meg Cabot X</td>
<td>77</td>
<td>Judith Krantz X</td>
</tr>
<tr>
<td>44</td>
<td>Agatha Christie X</td>
<td>78</td>
<td>Louis L’Amour X</td>
</tr>
<tr>
<td>45</td>
<td>Douglas Adamson</td>
<td>79</td>
<td>Robert Emery</td>
</tr>
<tr>
<td>46</td>
<td>Lauren Adamson</td>
<td>80</td>
<td>Frances Fincham</td>
</tr>
<tr>
<td>47</td>
<td>Eric Amsel</td>
<td>81</td>
<td>Timothy Findley X</td>
</tr>
<tr>
<td>48</td>
<td>V. C. Andrews X</td>
<td>82</td>
<td>Martin Ford</td>
</tr>
<tr>
<td>49</td>
<td>Isaac Asimov X</td>
<td>83</td>
<td>Ram Frost</td>
</tr>
<tr>
<td>50</td>
<td>Margaret Atwood X</td>
<td>84</td>
<td>Howard Gardner</td>
</tr>
<tr>
<td>51</td>
<td>Margaritia Azmitia</td>
<td>85</td>
<td>Stephen J. Gould X</td>
</tr>
<tr>
<td>52</td>
<td>Russell Banks X</td>
<td>86</td>
<td>Andrew Greeley X</td>
</tr>
<tr>
<td>53</td>
<td>Oscar Barbarin</td>
<td>87</td>
<td>Frank Gresham</td>
</tr>
<tr>
<td>54</td>
<td>Reuben Baron</td>
<td>88</td>
<td>John Grisham X</td>
</tr>
<tr>
<td>55</td>
<td>Gary Beauchamp</td>
<td>89</td>
<td>Steven Hawking X</td>
</tr>
<tr>
<td>56</td>
<td>Thomas Bever</td>
<td>90</td>
<td>Frank Herbert X</td>
</tr>
<tr>
<td>57</td>
<td>Brian Bigelow</td>
<td>91</td>
<td>S. E. Hinton X</td>
</tr>
<tr>
<td>58</td>
<td>Elliot Blass</td>
<td>92</td>
<td>Charles Hulme</td>
</tr>
<tr>
<td>59</td>
<td>Judy Blume X</td>
<td>93</td>
<td>Robert Inness</td>
</tr>
<tr>
<td>60</td>
<td>Dale Blyth</td>
<td>94</td>
<td>John Jakes X</td>
</tr>
<tr>
<td>61</td>
<td>Hilda Borko</td>
<td>95</td>
<td>Robert Jordon X</td>
</tr>
<tr>
<td>62</td>
<td>Ann Brasharest X</td>
<td>96</td>
<td>Frank Keil</td>
</tr>
<tr>
<td>63</td>
<td>Dan Brown X</td>
<td>97</td>
<td>Stephen King X</td>
</tr>
<tr>
<td>64</td>
<td>P.E. Bryant</td>
<td>98</td>
<td>Dean Koontz X</td>
</tr>
<tr>
<td>65</td>
<td>Meg Cabot X</td>
<td>99</td>
<td>Judith Krantz X</td>
</tr>
<tr>
<td>66</td>
<td>Agatha Christie X</td>
<td>100</td>
<td>Louis L’Amour X</td>
</tr>
</tbody>
</table>

174
<table>
<thead>
<tr>
<th>No.</th>
<th>Author Name</th>
<th>Age</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>Tom Clancy</td>
<td>57</td>
<td>John Steinbeck</td>
</tr>
<tr>
<td>24</td>
<td>James Clavell</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td><strong>Eoin Colfer</strong></td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Jackie Collins</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Max Coltheart</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>John Condry</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Stephen Coonts</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Carl Corter</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Diane Cuneo</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td><strong>Roald Dahl</strong></td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Geraldine Dawson</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Aimee Door</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Items with valid authors are noted with X.*

*New authors considered age-appropriate are indicated in bold.*
# APPENDIX C: WORKING MEMORY MEASURE

<table>
<thead>
<tr>
<th>Series Length</th>
<th>Trial</th>
<th>Item</th>
<th>Series Length</th>
<th>Trial</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>5,1</td>
<td></td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>8,7</td>
<td></td>
<td>2</td>
<td>6,8,9,2,4,5,7</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6,2</td>
<td></td>
<td>3</td>
<td>2,8,9,3,1,4,6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>9,3</td>
<td></td>
<td>4</td>
<td>7,6,3,5,9,4,1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>7,1</td>
<td></td>
<td>5</td>
<td>3,9,4,1,5,6,2</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>3,8</td>
<td></td>
<td>6</td>
<td>1,5,3,7,9,4,8</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3,2,6</td>
<td></td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9,2,7</td>
<td></td>
<td>2</td>
<td>6,8,9,2,4,5,7</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6,4,7</td>
<td></td>
<td>3</td>
<td>2,8,9,3,1,4,6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2,9,1</td>
<td></td>
<td>4</td>
<td>7,6,3,5,9,4,1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>7,5,3</td>
<td></td>
<td>5</td>
<td>3,9,4,1,5,6,2</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1,8,5</td>
<td></td>
<td>6</td>
<td>1,5,3,7,9,4,8</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3,8,5,4</td>
<td></td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2,7,9,1</td>
<td></td>
<td>2</td>
<td>3,6,7,5,2,8,4,1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>9,4,1,6</td>
<td></td>
<td>3</td>
<td>4,3,6,1,5,7,8,9</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6,3,9,5</td>
<td></td>
<td>4</td>
<td>3,8,5,4,2,7,9,1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>7,5,2,6</td>
<td>5</td>
<td>7,5,2,6,1,5,3,8</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>--------</td>
<td>---------</td>
<td>-------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>8,3,5,1</td>
<td>6</td>
<td>6,3,9,5,4,2,1,7</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>9,5,8,1,6</td>
<td>9</td>
<td>1,8,7,5,8,9,4,2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>8,3,6,2,9</td>
<td>2</td>
<td>8,5,7,4,1,6,7,2,3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4,1,8,5,3</td>
<td>3</td>
<td>1,8,5,3,2,6,7,8,4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3,7,5,8,2</td>
<td>4</td>
<td>3,9,7,4,2,1,5,8,6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1,4,7,9,5</td>
<td>5</td>
<td>6,2,9,8,7,4,3,5,1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>6,2,9,5,8</td>
<td>6</td>
<td>1,5,3,7,9,8,4,6,2</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX D: NOVEL WORD PASSAGES AND SCORING SCHEME

First mention of novel words are indicated in italics. Defining pieces of information are indicated in bold. All text was presented in normal font.

Text 1

Heather and her mother were headed home Friday night. As they made the turn onto the road they lived, they saw a lortin on the side of the road. It was hard to see at first because its colour blended with the trees bordering the side of the road. Heather’s mom stopped the car so that the animal wouldn’t be frightened. They looked at each other and agreed that they weren’t sure what kind of lortin it was. Her mother turned to Heather, and said, “Maybe we’re wrong Heather. It could be a horse.” Heather shook her head and replied, “I can see how the size would make you think that, but it can’t be mom. Look! Its tail is much too short.” Before they could look any longer, the lortin disappeared into the forest.

Scoring. Maximum score for lortin was three points. One point awarded for each defining piece of information: animal, as big as a horse, has a short tail. The following examples scored 2 points: animal and as big as a horse, similar to a horse, a large animal resembling a horse, an animal similar to a horse, horse-like, a large animal. Specific mention of the size of the tail (e.g., shorter than a horses, or a short tail) was required to score the third point.

Text 2

Jake wanted to buy a present for his sister for Christmas. He decided to buy her a palmit. As he looked through the store, he saw many boxes that could hold her things, but nothing
seemed to come close to what he wanted to buy for his sister. Turning down the next aisle, he saw what looked to be a palmit. He wound the switch at the bottom of it, and listened to the music that played. He knew his sister would like the tune. Jake became excited that he had found a palmit, but then he realized something was missing. He asked the salesgirl if she had one that locked. The salesgirl said, “Sorry, I think you are looking for a palmit, and I just sold the last one.” Jake decided to buy the box he had been looking at for his sister, even though it wasn’t really a palmit.

**Scoring.** Maximum score for *palmit* was three points. One point awarded for each defining piece of information: box, musical, locable. The following examples scored 2 points: a music box, a musical box, a box that plays music. One point was scored for the following: a music dcevice, a music player, a toy box, a jewellery box (with toy or jewellery considered as an unnecessary elaboration. Zero points scored for jukebox or switch at bottom. Specific mention of the size of the tail (e.g., shorter than a horses, or a short tail) was required to score the third point.

**Text 3**

Mary Smith was looking to move. She told her real estate agent that she wanted a *slond*. The real estate agent said, “You mean something that is *large and spacious*?” Mary said “Yes”, and they started to look at houses. The first house was nice, but Mary didn’t like the neighbourhood. When she walked into the second house, Mary said, “But this *isn’t a slond – it has a second floor!*” The real estate agent apologized and took her to see several more houses. In the next house, Mary said, “I like it, but all I can see out the windows is farmland.” In the last house, Mary said, “Finally a real slond – those *windows have a great view of the ocean!*”
Scoring. Maximum score for slond was four points. One point awarded for each defining piece of information: a house, one floor, large and spacious, by the ocean (or, view of the ocean). Two points were awarded for mention of bungalow. The following example scored one point: somewhat like a cottage, resembling a condominium, home. Zero points were awarded for a view, no floor, or big windows.

Text 4

Marion and Sam had spent their vacation last summer by the ocean. They had really enjoyed themselves because they loved the smell of the salt air and availability of fresh seafood. This morning Marion went down to the beach and filled up a jar with baby clams that the tide had washed up, and put it in the basket she had brought with her. She would need them to make cistel tonight. Marion felt that their vacation was going by too quickly. Sam and Marion had taken turns preparing different food dishes for each other for dinner. Cistel was Marion’s favorite way of making pasta sauce. As she went by the market she bought garlic, basil and olive oil. She wouldn’t need the basil to make tonight’s dinner, but cistel had to have the other things she had in her basket. As she headed home, she met Sam at the door. He peered into the basket Marion carried, and let out a big smile.

Scoring. Maximum score for cistel was three points. One point awarded for each defining piece of information: food (pasta sauce, or sauce), clams (baby clams, or seafood, garlic and olive oil (.5 mark was awarded if only 1 item mentioned). Two points were awarded for mention of seafood sauce.

Text 5
In the 1600s, *Minalds* were a common sight on the oceans. Today they are mostly seen in naval museums. They are often confused with other *ships* made during the same period, but the *ornate carvings* which grace their decks make them unique. Recently, at a private auction, one person was prepared to pay five million dollars for what he thought to be a Minald. It eventually sold for only half that price once it was found out that the wood used to make the deck floor was made of teak. By the early 1700s, the *cherry wood* needed in the building of Minalds had become too costly, and their production came to an end. (117 words)

**Scoring.** Maximum score for *Minalds* was three points. One point awarded for each defining piece of information: ship, ornate carvings on deck, made of cherry wood. Unnecessary details or embellishments (including seen oceans, seen in naval museums, old, from 1600s, confused with other ships, not made of teak, production came to an end, goes on water, stopped being made, expensive) were not awarded points

**Text 6**

People from different parts of the world celebrate local traditions. *Diboll* happens only in *Southern Austria*. People who choose to participate in Diboll know that they must meet at midnight at the foot-path at the bottom of Mount St.Veit. Both young and old people of the villages come prepared with their torches lit, and their knapsacks filled with food and drink. Not all of them will meet the goal of *climbing over two mountains* by daybreak. Given the time and energy needed for Diboll, it is not surprising that it only *happens once a year*. Some people say that the successful completion of Diboll even once in your life means that you will go straight to heaven. Other people feel that it is simply a way of enjoying time with your friends and family. (136 words)
Scoring. Maximum score for Diboll was four points. One point awarded for each defining piece of information: yearly, Southern Austria (or Austria), people climb over mountains, from mid-night to daybreak; happens during night. One and half points were awarded for climbed over the mountains by daybreak with no mention that the start was at midnight or if the start at midnight. Unnecessary information or embellishments (including celebration, tradition, fun with family and friends, or a religious interpretation of the meaning i.e. going straight to heaven, bringing torches and knapsacks filled with food) were not awarded points.
APPENDIX E: SUMMARY WRITING TEXT (“CONTROLLING PESTS”)

Animal and insect pests have bothered people for thousands of years. One way to control these pests is through the use of chemicals called pesticides. However, pesticides may damage the environment and cause a number of related problems.

Some time ago in Malaysia, there was a dangerous type of mosquito. These mosquitoes were spreading a serious disease called malaria. Malaria is a disease that attacks blood cells. It causes chills, high fever, and can even result in death.

American scientists thought that they could solve this problem. They decided to try to kill the mosquitoes with a pesticide called DDT. The DDT was sprayed in the forests and swamps where the mosquitoes were found. Small planes were used for the spraying to make sure that large areas of land were covered. The pesticide was very effective in controlling the mosquitoes and the spread of malaria.

However, other problems began to occur. The poisonous DDT quickly got into the fresh water system. This water could not be used by people or animals. Many useful insects and animals became sick and died. For example, a large number of cats died after eating tree-climbing lizards. The lizards had poisoned by the DDT. The loss of the cats caused a huge increase in the number of rats. The rats carried fleas. The fleas spread a disease called the bubonic plague. Many people became sick once again.
The loss of the tree-climbing lizards caused other problems. The Malaysian people lived in huts made of straw. There were insects that ate the straw, but they were controlled by the lizards. When the lizards died, the huts became infested with insects and collapsed.

Today, scientists want people to stop using chemicals to kill pests. Instead, they suggest that we use natural enemies to control the pests. All pests have natural enemies. In many cases, the enemy can be used to control the pest population.

That’s what they tried in Australia in the 1930’s. The Australian sugar cane farmers were having a problem with grey-backed beetles. The beetles were eating the young cane shoots. As a result, the entire sugar cane crop was threatened.

Then Australian scientists found a solution. They discovered that the natural enemy of the grey-backed beetle is the giant toad. Giant toads live in South America. The scientists brought hundreds of the giant toads to Australia. They turned them loose in the sugar cane fields. The toads showed a great appetite for grey-backed beetles and other insect pests. They ate so many of the beetles that within a few years, the threat to the sugar cane crop was over. Because the toads had saved the sugar cane crop, they became known as cane toads.

However, as the number of grey-backed beetles declined, the number of giant toads grew. The toads moved out of the cane fields in search of food. They began eating many useful insects such as honeybees. Ever since the coming of the cane toads, Australian beekeepers have suffered great losses. The sugar cane problem was solved, but now there was a cane toad problem.

In the warm coastal areas of northeastern Australia there used to be large numbers of red-bellied black snakes. The snakes were useful because they helped to control the mice and rats. When the toads came along the snakes ate them. Unfortunately for the snakes, cane toads have
two poison glands behind their head. An adult toad packs enough poison to kill almost any animal that eats it. The red-bellied black snakes died when they ate the cane toads.

Many farmers in Australia hate cane toads. Before the toads came, the red-bellied black snakes controlled the mice and rats. When the toads came and the snakes died, the mice and rats increased in number and ate all the farmers’ grain. Many farmers had to give up their farms because they went bankrupt.

Pesticides cause problems because they are not a natural way of controlling pests. But, the cane toads caused problems too. Is it wrong to use natural controls? Were the cane toads really natural?
APPENDIX F: PROPOSITIONAL TEMPLATE AND SCORING METHOD FOR SUMMARY WRITING TASK

A propositional template (see Table F.1) was developed for the scoring of the “Controlling Pests” summary writing task.

**Micropropositions**

All important micropropositions (M1) and unimportant micropropositions (M0) were identified directly from the propositional analysis of the text (see Table F.1). Thus, all M0 and M1 propositions could be selected from the text and required no integration, generalization, or construction processes. The distinction between M1 and M0 ideas was based on the relevance of the microproposition, and the contribution it made in supporting higher level ideas. Whereas M1 micropropositions provided direct support for a higher level idea, M0 micropropositions provided little or no support for higher level ideas. There were a total of 24 level M1 ideas, and 28 level M0 ideas. M1 micropropositions were awarded one point. M0 micropropositions were not awarded any points.

**Macropropositions**

The “Controlling Pests” text comprised 12 macropropositions (M2) representing the main ideas of each paragraph and are also identified in Table F.1. The macropropositions were of two types: directly-stated and constructed macropropositions.

**Directly-stated macropropositions.** Five of these macropropositions (M2-1, M2-2, M2-4, M2-7, M2-8) were directly stated in the text. One point was awarded for each of the main
ideas whether it was expressed in verbatim or paraphrase form. An example of a directly stated M2 proposition and two paraphrased versions that were accepted for this proposition follow.

M2-7 Instead, they suggest that we use natural enemies to control the pests.

1. Scientists say that farmers should use a more natural way of getting rid of pests. All pests have a natural enemy.
2. Scientists decided to use natural pest control.

**Constructed macropropositions.** The remaining 7 macropropositions (M2-3, M2-5, M2-6, M2-9, M2-10, M2-11, M2-12) were constructed through the integration of two or more directly stated propositions. Constructed propositions included those concerning the effects and consequences of the pest, the use of pesticides and the natural controls. To receive credit for these macropropositions, the summary had to include either two examples of effects or consequences, or one example with a general statement discussing the problems that were caused. Each example that was provided also scored one point at the micropropositional (M1) level. The following excerpts from student summaries illustrate this scoring format.

M2-5 The negative effects of using DDT included poisoned water and the loss of useful animals and insects.

1. The chemical poisoned the water and killed the animals that drank it.
2. DDT got into the water system and many animals died from the poison.
3. …but the DDT had side effects which caused the deaths of cats and other useful animals.
4. DDT did not work because the poison got into the fresh water system which is used by people and animals.
If only a single example (see examples 5 and 6 below) of an effect or consequence was included, one point for the lower level micro-proposition (M1) was awarded.

5. DDT was sprayed and it got into the water.

6. Pesticides poison the fresh water system.

**Thematic Propositions**

Four thematic propositions (M3) were identified from the text and are also presented in Table F-1. In contrast to the M2 level macropropositions, that represented main ideas derived from each paragraph, the thematic propositions reflected a more abstracted generalization and integration of the main ideas of the text in general. As well, a thematic proposition requires the integration of information in the text with prior knowledge about the topic. For example, the thematic propositions stating that *both pesticides and natural enemy methods of pest control create an upset in the balance of nature* requires the integration of the important information presented in the text, as well the relation of this information with prior knowledge about ecological systems and the delicate balance of nature. Students were awarded one point for each thematic proposition that was identified in their summary. In many cases, the propositions were not stated as succinctly as they were for the template. However, if the summary contained a sentence or sentences that conveyed one of these thematic elements one point was awarded. The following excerpts provide two examples of each proposition receiving a one point score.

**M3-1 Overview of methods (pesticides and natural controls) used to control pests.**

1. Natural enemies or pesticides, what should we use?
2. This article gives us a brief glimpse of pest control. The two ways that pest control are implemented are either chemicals that are manmade or introducing a natural predator to eliminate the pest.

M3-2 Both pesticides and natural controls cause an upset in the balance of nature

1. So, there really is no safe way of getting rid of pests, without throwing everything else off balance.

2. …screws up the food chain and ecosystem.

M3-3 Pesticides and natural controls are similar in that they both solve the pest problem, but also cause environmental damage and related problems.

1. So if you have a problem either way the solution causes a problem.

2. …gets rid of pest but also changes the environment in drastic ways.

M3-4 New ways are required to solve animal and insect pest problem, or recognition that controlling pests is a difficult problem.

1. So I guess they need to find a new way to deal with pests because pesticides and natural enemies don’t work.

2. Controlling pests is difficult.
Propositional Template for Summary Writing Task

<table>
<thead>
<tr>
<th>Level</th>
<th>No.</th>
<th>Proposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2</td>
<td>1</td>
<td>Animal and insect pests have bothered people for thousands of years.</td>
</tr>
<tr>
<td>M1</td>
<td>1</td>
<td>One way to control these pests is through the use of chemicals called pesticides.</td>
</tr>
<tr>
<td>M2</td>
<td>2</td>
<td>However, pesticides may damage the environment and cause related problems.</td>
</tr>
<tr>
<td>M0</td>
<td>2</td>
<td>Some time ago</td>
</tr>
<tr>
<td>M1</td>
<td>2</td>
<td>in Malaysia,</td>
</tr>
<tr>
<td>M2</td>
<td>3</td>
<td>there was a dangerous type of mosquito. These mosquitoes were spreading a serious disease.</td>
</tr>
<tr>
<td>M1</td>
<td>3</td>
<td>called malaria.</td>
</tr>
<tr>
<td>M0</td>
<td>3</td>
<td>Malaria is a disease that attacks the blood cells.</td>
</tr>
<tr>
<td>M0</td>
<td>4</td>
<td>It causes chills, high fever, and can even result in death. (For example, there was a problem with mosquitoes).</td>
</tr>
<tr>
<td>M0</td>
<td>5</td>
<td>American scientists thought they could solve this problem.</td>
</tr>
<tr>
<td>M2</td>
<td>4*</td>
<td>They decided to try to kill the mosquitoes with a pesticide called DDT.</td>
</tr>
<tr>
<td>M1</td>
<td>4</td>
<td>The DDT was sprayed</td>
</tr>
<tr>
<td>M0</td>
<td>6</td>
<td>in the forests and swamps where the mosquitoes were found.</td>
</tr>
<tr>
<td>M0</td>
<td>7</td>
<td>Small planes were used for the spraying</td>
</tr>
<tr>
<td>M0</td>
<td>8</td>
<td>to make sure that large areas of land were covered.</td>
</tr>
</tbody>
</table>
The pesticide was very effective in controlling the mosquitoes and the spread of malaria.

However, other problems began to occur.

The poisonous DDT quickly got into the fresh water system.

This water could not be used by people or animals.

Many useful insects and animals became sick and died.

For example, a large number of cats sided after eating tree-climbing lizards.

The lizards had been poisoned by the DDT.

The loss of the cats caused a huge increase in the number of rats. The rats carried fleas. The fleas spread a disease called the bubonic plague.

Many people became sick once again

(The effects of DDT included the poisoning of water and killing of useful insects and animals.)

The loss of the tree-climbing lizards caused other problems.

The Malaysian people lived in huts made of straw.

There were insects that ate the straw but they were controlled by the lizards.

When the lizards died, the huts became infested with insects and collapsed.

(Further consequences of the DDT included the loss of many animals, the spread of new diseases, and the loss of homes.)

Today, scientists want people to stop using chemicals.

Instead, they suggest that we use natural enemies to control the pests.
All pests have natural enemies.

In many cases, the enemy can be used to control the pest population.

That's what they tried in Australia in the 1930's.

The Australian sugar cane farmers were having a problem with grey-backed beetles.

The beetles were eating the young cane shoots.

As a result, the entire sugar cane crop was threatened.

Then Australian scientists found a solution.

They discovered that the natural enemy of the grey-backed beetle is the giant toad.

Giant toads live in South America.

The scientists brought hundreds of toads to Australia.

They turned them loose in the sugar cane fields.

The toads showed a great appetite for grey-backed beetles and other insect pests.

They ate so many of the beetles that within a few years, the threat to the sugar cane crop was over.

Because the toads had saved the sugar can crop, they became known as cane toads.

(The toads solved the beetle problem.)
However, as the number of grey-backed beetles declined, the number of giant toads grew.
September 7th, 2006

LETTER OF INFORMATION

Understanding Reading Comprehension in Adolescent Students

I am writing to request your participation in a research project about adolescent reading comprehension. I am a teacher at St. Theresa’s High School, and am now on leave while I pursue studies at the Faculty of Education at Queen’s University in Kingston. This research has been cleared by the ethical review committee of Queen’s University, by your school board and classroom teacher. Please read the information below. If you agree, please sign the attached sheet (“Consent Form”), take it home for your parent or guardian to sign and then return it to your classroom teacher.

Some background

The ability to read and to understand the meaning from what is read is a skill that is important to all forms of personal learning, intellectual growth and success in school. Results from the Ontario Secondary School Literacy test show that a substantial number of students have not achieved adequate levels of reading comprehension by Grade 10. While substantial progress has been made in understanding learning to read in young children, far less is known about reading in adolescents, and about comprehension (i.e. reading to learn).

Reading comprehension is a very complex process. It depends upon a number of skills such as word reading, verbal ability, inferencing, and comprehension monitoring to name a few. Unfortunately, the government literacy tests do not provide this kind of detailed information. Additionally, it is uncertain how well the EQAO literacy tests that you completed in earlier grades may have been used to predict your performance in later grades. To improve reading comprehension achievement, it is necessary to understand the factors and underlying processes involved so that the best type of instruction can be designed.

About the project

This project will use measures that research has shown to be important for effective reading comprehension. There is considerable evidence that shows that reading comprehension is held back by inefficient word readings and by the processes known to underlie word reading in younger readers. However, little research has been carried out with adolescent readers to determine if this, too, may be a source of difficulty. There is some suggestion that some adolescents can get around their word-reading problems and still attain adequate reading comprehension. Therefore, while efficient word reading is important, it is insufficient for successful comprehension to occur. Comprehension problems can occur because of weakness in the ability to carry out inferences, to integrate
information found in the text with the reader’s background knowledge, and inadequate comprehension monitoring skills to name a few.

**How you will be involved**

If you agree to participate, I ask that both you and your parent or guardian sign the consent form attached on the last page. I will arrange with your teacher convenient times to visit the school. You and your classmates will take part in a number of activities that have been selected to assess particular aspects considered to be important to reading comprehension. These will include activities such as summary writing, word reading, word meanings, identifying the main idea, inferencing and integration, comprehension monitoring, as well as questionnaires about how actively you are engaged in reading.

These activities will be administered over three class periods (60 minutes each). Two of the sessions will occur in class as a group, and one session will involve you meeting with me individually for one period in a quiet place in the school. Some of these activities will involve a computer, while others will be paper and pencil activities. Each session will only begin if you agree to take part. If you are away on a day when the class is taking part in an activity, I will make arrangements for you to complete the activities at another time. You may decide to end your participation in the study at any time. The activities will take place from September to December. I will do my best to ensure that you enjoy the activities. If you lose interest, I will end the session and continue another day. Your classroom teacher has agreed for you to participate in these sessions. Your class grade will not be affected by your participation.

I also ask your permission to obtain detailed results from your Grade 3, Grade 6, and the Grade 10 (when they become available) EQAO literacy results. This will require access to the identification number assigned to you by EQAO on each of these assessments. These numbers are in your Ontario School Record (OSR) available in Student Services. Finally, I ask your permission to check your OSR to determine the nature of any I.E.Ps you may have, or may have had in the past.

When the study is completed, and the information has been analyzed, a summary of the results will be given to you, if you wish.

**What we promise**

The research team consists of: Dr. John Kirby (Professor), Dr. Elizabeth Lee (Associate Professor) (Faculty of Education) and Dr. Brian Butler (Department of Psychology) at Queen’s University and myself (doctoral student in Education at Queen’s University). The information you supply may result in publications of various types, including journal articles, professional publications, newsletters, books, reports and instructional materials for teachers. We promise that the information you supply will be kept confidential, and that no individual information about you, or the school will ever be published or made available to anyone outside the research team. Your name will only appear on raw data sheets, which will be kept indefinitely in a secure cabinet in a locked room at Queen’s University; only code numbers will appear in any other file. There are no foreseeable risks or discomfort associated with any of these measures. Participation
is voluntary. You have the right to not complete any measure you do not want to complete, and to withdraw from the study at any stage; withdrawal, or a decision to not participate in the study, will have no effect upon your grades.

Thank you very much for considering this request. If you have any questions about the project, please contact Bo White at (705-549 6595) or Dr. John Kirby at the Faculty of Education, Queen’s University (613-533-6220). For questions, concerns or complaints about the research ethics of this study, contact the Dean of Faculty of Education, Dr. Rosa Bruno-Jofré, (613 533-6081; brunojr@educ.queensu.ca), or the chair of the General Research Ethics Board, Dr. Joan Stevenson, (613 533-6081; stevensj@post.queensu.ca).

Sincerely,

Bo White
Ph.D. Candidate
STUDENT CONSENT FORM

“Understanding Reading Comprehension of Adolescent Students.”

I understand that I am being asked to participate in a research project entitled *Understanding Reading Comprehension in Adolescent Students*. I understand that the purposes of the study are to investigate the reading and reading comprehension skills of adolescent students, to relate these to the scores obtained on the Grade 10 Ontario Secondary School Literacy Test (OSSLT), and to determine the extent to which the scores on the Grade 10 OSSLT are predictable from earlier Ontario Ministry of Education assessments (EQAO grades 3 and 6). I have read, understood and retained a copy of the Letter of Information, and I have had any questions answered to my satisfaction.

I understand that my participation in September – December 2006 will be administered over three sessions each taking one classroom period (70 minutes each). I will be asked to complete a variety of reading and spelling measures. I will also be asked to complete two questionnaires related to my reading and learning experiences. Two of these sessions will be carried out in class as a group, and the remaining session be individually administered. My answers will be recorded on paper or by a computer.

I understand that all data will be kept confidential and in locked rooms, that my teacher will not have access to any individual scores, and that no individual will be identified in any publication. I can withdraw from the study at any time without any consequence. I understand that there are no known risks, discomforts or inconveniences involved in the study. Participation in the study and performance on the tests will have no effect upon my grades.

I am aware that I can contact the researcher, Bo White by telephone at 705 549 6595 (whiteb@educ.queensu.ca), or her supervisor Dr John Kirby at Queen’s University, Faculty of Education at 613-533-6000 ext. 36220. I am also aware that for questions, concerns, or complaints about the research ethics of this study, I may contact the Dean of the Faculty of Education, Dr. Rosa Bruno-Jofré, 613-533-6210, (brunojor@educ.queensu.ca) or the chair of the General Research Ethics Board, Dr. Joan Stevenson 613-533-6081 (stevensj@post.queensu.ca).

Please initial all that apply below and sign the following and return it to me. I will be available to collect the permission forms during class.

☐ I consent to participate in the *Understanding Reading Comprehension in Adolescent Students* project.

☐ I give permission for the researcher to access my Ontario School Records to obtain my EQAO identification numbers from the literacy tests I may have written in Grades 3, 6 and will write in grade 10.

☐ I give permission for the research to access my EQAO literacy test scores that I may have written in Grades 3, 6 and will write in grade 10.

☐ I do not have an Individual Education Program (IEP).

☐ I have an Individual Education Program (IEP), and give permission for the researcher to access my Ontario School Records to determine the nature of my identification.
PARENT/GUARDIAN CONSENT FORM
“Understanding Reading Comprehension of Adolescent Students.”

I understand that I am being asked to agree to have my son/daughter participate in a research project entitled Understanding Reading Comprehension in Adolescent Students. I understand that the purposes of the study are to investigate the reading and reading comprehension skills of adolescent students, to relate these to the scores obtained on the Grade 10 Ontario Secondary School Literacy Test (OSSLT), and to determine the extent to which the scores on the Grade 10 OSSLT are predictable from earlier Ontario Ministry of Education assessments (EQAO grades 3 and 6). I have read, understood and retained a copy of the Letter of Information, and I have had any questions answered to my satisfaction.

I understand that my son’s/daughter’s participation in September – December 2006 will be administered over three sessions each taking one classroom period (70 minutes each). My son/daughter will be asked to complete a variety of reading and spelling measures. He/she will also be asked to complete two questionnaires related to their reading and learning experiences. Two of these sessions will be carried out in class as a group, and the remaining session be individually administered. Their answers will be recorded on paper or by a computer.

I understand that all data will be kept confidential and in locked rooms, that my son’s/daughter’s teacher will not have access to any individual scores, and that no individual will be identified in any publication. My son/daughter can withdraw from the study at any time without any consequence. I understand that there are no known risks, discomforts or inconveniences involved in the study. Participation in the study and performance on the tests will have no effect upon my son’s/daughter’s grades.

I am aware that I can contact the researcher, Bo White by telephone at 705 549 6595 (whiteb@educ.queensu.ca), or her supervisor Dr John Kirby at Queen’s University, Faculty of Education at 613-533-6000 ext. 36220. I am also aware that for questions, concerns, or complaints about the research ethics of this study, I may contact the Dean of the Faculty of Education, Dr. Rosa Bruno-Jofré, 613-533-6210, (brunojr@educ.queensu.ca) or the chair of the General Research Ethics Board, Dr. Joan Stevenson 613-533-6081 (stevensj@post.queensu.ca).

Please initial all that apply below and sign the following and have your son/daughter return it to me. I will be available to collect the permission forms during class.

☐ I consent to allow my son/daughter to participate in the Understanding Reading Comprehension in Adolescent Students project.

☐ I give permission for the researcher to access my son’s/daughter’s Ontario School Records to obtain my EQAO identification numbers from the literacy tests he/she may have written in Grades 3, 6 and will write in grade 10.

☐ I give permission for the research to access my son’s/daughter’s EQAO literacy test scores that he/she may have written in Grades 3, 6 and will write in grade 10.

☐ My son/daughter does not have an Individual Education Program (IEP).
☐ My son/daughter has an Individual Education Program (IEP), and I give permission for the researcher to access his/her Ontario School Records to determine the nature of his/her identification.

Child’s Name: _______________________________  Date: ______________________

Parent’s Name: _______________________________  Signature: ____________________

Please include your email or postal address at the bottom of this sheet, if you wish to receive a copy of the results of this study.