This is to certify that the

dissertation entitled

The Effect of Item Text Characteristics

on Children's Growth in Reading

presented by

Hye-Sook Park

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Educational Psychology

P. D. Pearson Ralph Putnam

Major professor

Date Dec 18 1998

The Effect of Item Text Characteristics on Children's Growth in Reading

By

Hye-Sook Park

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational Psychology
and Special Education

1999

ABSTRACT

The Effect of Item Text Characteristics on Children's Growth in Reading

By

Hye-Sook Park

This study investigates children's growth in reading reflected on the Peabody Individual Achievement Test (PIAT) reading comprehension item responses from the National Longitudinal Survey of Youth data over several years. Based on the idea that reading comprehension is determined by characteristics of both readers and texts, this study investigates the relative impact of both. Using a three-level hierarchical generalized linear model, in which items (level-1) are nested within time points (level-2) and time points are nested within individuals (level-3), this study assesses relationships among text characteristics, cognitive abilities, environmental factors, and reading ability (as indexed by the Peabody text).

Reading ability did not grow at a constant rate; in fact it exhibited variable patterns that were influenced by verbal memory and text characteristics in different ways at different points in children's reading development. In general, short sentences, frequently used vocabulary, and high density facilitated reading comprehension, but the temporal influences of the patterns of three text characteristics differed.

The effect of age on children's reading comprehension was manifested differentially depending upon sentence characteristics. In the case of sentence length, the effect of age was manifested only with short sentences. The positive contribution that frequently used vocabulary made to reading comprehension increased over years, but the

growth rates were also different. The effect of age on reading comprehension was greater with sentences written using high frequency vocabulary than with low frequency vocabulary. The effect of propositional density increases constantly. The effect of age on reading comprehension was manifested greatly with high density sentences, that is, coherent sentences, rather than with low density sentences.

In addition, verbal memory was statistically significant in predicting both the average effect of sentence length over time and the rate of growth of sentence length slope. There was an interaction effect between verbal memory and length of sentences over time. In the case of short sentences, the effect of verbal memory was practically as well as statistically significant. However, in the case of long sentences, the effect of verbal memory was almost absent. As verbal memory increased, vocabulary frequency had a greater effect on reading ability. However, verbal memory did not influence the effect of propositional density.

The differential contribution of each psycholinguistic variable over time implies that achievement, as measured by a reading comprehension test, is a complex entity that is greatly dependent on the nature of the text contained in the test.

To the Almighty

who made this possible

ACKNOWLEDGEMENTS

Although my dissertation research was conducted based on information processing theory due to the affordances/constraints of the outcome measure, my work acknowledges that knowledge is socially constructed. I owe directly and indirectly so many things to professors and friends who have made this work possible. Looking backwards, due to their support, I feel that the period of conducting this research was the happiest one in my graduate program.

I was blessed to work with my committee members, who are the model educators and researchers that I have admired. As my dissertation director, Professor P. David Pearson provided indescribable mentoring and support. He provided an intellectual niche for me and provided me with different types of financial support. In spite of his overwhelmingly busy schedule, I could get feedback from him whenever I needed it. His accessibility to me both via e-mail and face-to-face interaction within 24 hours facilitated my progress. His father-like support has sustained my energy, spirit, and enthusiasm for learning.

I also would like to express my deepest thanks to Professor Stephen W. Raudenbush, the great motivator in inspiring my interests and creating potentials as a researcher in quantitative research. Because of him, I came to like statistics. He has provided valuable feedback on each phase of my dissertation, even as he was moving to another university. His warm and understanding smiles have also given me the energy to go over work again and again during various stages of my graduate program. He always told me to do "ground-breaking work," and at the same time he did not forget to tell me to enjoy the Michigan sunshine.

Sincere thanks also goes to Professors Ralph T. Putnam, my advisor at the last stage of my graduate program, for setting high standards on my study, listening to me, editing my papers, and supporting me in many situations. I have known Professor Thomas J. Luster since the first year during my master's program. He informed me about the National Longitudinal Survey of Youth data for this research. His on-going kindness and warm support have made my life at Michigan State University easy.

Special thanks goes to Michael C. Rodriguez, for his kindness and timely help. I learned something about measurement from him, too. I am also grateful to Professor Barbara K. Abbott because of her kindness when I felt ambiguity in analyzing propositions. I also would like to say thank you to Dr. Randall Fotiu for some technological support at the beginning of my study, Yasuo Miyazaki and other HLM classmates for posing intellectual challenges to me and providing feedback on my study; Professors William Mehrens, Betsy Jane Becker, and Kenneth Frank for their kindness in giving me the opportunities to learn more about measurement and statistics; and Professor Stephen L. Yelon for his advice and kindness during my wandering periods. My thanks extends to Professors Richard Prawat, John Schwille, and James Gavelek, my former advisor, for providing me with direct and indirect financial support and encouragement.

Thanks goes to friends, Rena and Robert Atherton, my host family; Hai-young and Dr. Sung-Jun Kim, Hee-Sook Park, Ok-Sook Park, Mr. Yoon-seng Kim, Sheila Moore, Ailing Kong, Drs. In-Kyung Lee, Kedmon Hungwe, Myung-Ae Bang, and Leticia and Rodolfo Altamirano for unforgettable friendships, help, and prayers for me especially during my physical ailment. Thanks goes to my office-mates and Gary Affholter for their support. Thanks to Professors Shin-You Lee and Sang-Ok Park at Hong-Ik University, Dr. N. M. Mohan Pankaj at Australia National University, Mr. Jae-Hyuk Yang, Shinwon Middle School principal, for their encouragement, and my student, Ju-sub Yoon for reminding me that I am realizing my dream.

Finally, my thanks goes to my family members. To my parents, Keum-Nam Park and Young-Su Kim, for their high expectation on my education and their sacrificial support for their children's education; to my sister, Kyung-Sook Park for her warm support and cheering; to my younger brother and his wife, Jae-Hyun Park and Keum-mi Lee, especially to their children, Ji-Su and Sang-Su for their smiles and cheering, and to my older brother, Kyung-Sun Park who has indirectly led me to become a persistent learner. I also would like to extend my thanks to many other friends and faculty members who have helped me grow as a researcher at Michigan State University.

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

For decades, studies on readability have been conducted to understand the effect of text characteristics on reading comprehension. However, no studies have been conducted to investigate how the effect of text characteristics on reading comprehension changes as children grow older.

This study investigates how the linguistic characteristics of text interact with characteristics that children bring to the classroom either by virtue of nature or experience. This study explores factors that explain or account for the growth in beginning readers' abilities at ages 6, 8, and 10 in terms of potentially explanatory variables: (a) psycholinguistic variables such as sentence length, word frequency, and idea density; (b) changing home environmental factors; and (c) time invariant individual characteristics such as race, gender, verbal memory, and testing time. In addition, this study investigates how individual characteristics interact with psycholinguistic variables in explaining growth in reading.

Reading achievement was measured by the Peabody Individual Achievement Test (PIAT) Reading Comprehension items across three time points over four years as a part of the National Longitudinal Study of Youth (NLSY). These data are the primary outcome measures for this investigation.

It is commonly believed that reading comprehension is determined by the joint influence of the characteristics of readers and the texts they read, with the assumption that ability is itself the joint effect of biological (genetic) and environmental factors. The present study builds on a long tradition of readability studies in the sense that it incorporates text characteristics in the model.

Traditionally, studies on readability have used regression models to explain the difficulties of texts. Some of these studies put linguistic and psycholinguistic factors into the model to explain text difficulties. Early readability studies (Chall et al., 1948; Flesch, 1943) investigated only observable text characteristics (e.g., number of words in a sentence, number of syllables in a word, number of prepositions, and vocabulary frequencies). More recent studies have tried to explain text difficulties by incorporating reader factors, such as reader's prose-processing capability (Kintsch, 1979). Carver (1977) and Stenner (1997) measured both the difficulties of texts and the ability of readers by attempting to place the two constructs on the same scale. However, in spite of these researchers' contributions to the area of reading comprehension, questions still remain regarding how the importance of text characteristics differs with respect to readers' abilities. Text characteristics interact with the characteristics of readers, and readers' abilities may influence the perception of text characteristics. Thus, it is important to investigate the changing patterns of influence of linguistic and psycholinguistic variations of texts, especially as they are moderated by changes in children's underlying reading abilities and cognitive growth. Based on information processing theory, the study will investigate the NLSY children's PIAT reading comprehension item responses using hierarchical generalized linear models (HGLM).

The NLSY PIAT reading comprehension item responses provide important information for understanding beginning readers' development from ages 6 to 10. According to Chall's (1983) reading development scheme, children undergo six different reading stages before reaching adulthood: pre-reading, initial decoding, reading for confirmation of knowledge, reading for obtaining conventional knowledge, reading with multiple view points, and the construction/reconstruction of knowledge. However, no studies have investigated the changing impact of text difficulty as children progress from one stage to the next. In addition, no studies have investigated how the text characteristics interact with children's individual characteristics. Especially rare is the use of item responses by the same subjects to the same test across several time points over several years, as is the case in this study. This longitudinal perspective will help us examine the complex array of factors that influence reading development more extensively and more accurately. The use of a common metric and a single group of subjects across years eliminates the confounding that might occur if either a different assessment instrument were to be employed across time or different subjects were incorporated at each time point. In addition, this study will avoid some problems commonly found in this sort of research: If only two time points are used, it is difficult to assess the trends (growth) of children's reading development across years; cross-sectional designs obscure the assessment of the individual children's development across years due to cohort effects.

To explain the immediate text processing phenomena at each time point, this study will be based on information processing theory. In fact, the very structure of the PIAT suggests a grounding in information processing theory. The characteristics of the PIAT items, procedures, and underlying assumptions about reading processes are consistent with information processing theory.

The PIAT reading comprehension test comprises 66 items, each a single sentence. As the test progresses, sentences get longer and the words used become less common. The PIAT reading comprehension test is conducted by asking children to read a sentence only once, turn a page, and then to select one of four pictures that describes the sentence. The PIAT uses a range-finding approach to item selection for each individual by giving a certain range of items that is appropriate to readers' ability levels based on the PIAT reading recognition score (a word identification test). A basal level (a range of easy items) and a ceiling level (a range of very hard items) are found for each child, and a final score for any individual is based upon performance on these items that fall between the basal and ceiling levels.

The assumption of reading found in the PIAT test is that comprehension is a process of finding meaning in a text. The meaning of the text exists independently of the reader, since the reader has to choose the one correct meaning out of four options. When children take the test they must engage recall (Carroll, 1972), or short-term memory. (They turn the page after reading the sentence in order to see the four picture choice.) Carroll (1972) argued that having readers answer questions without the text present overemphasizes the memory component rather than measuring pure comprehension of text reflected in lexical knowledge, grammatical knowledge, and an ability to locate facts in a paragraph.

As suggested, this test also requires the child to invoke short-term memory (STM) or working memory. As indicated by Jorm (1983) and Morrison, Giordani, and Nagy (1977), there exists a relationship between reading ability and STM. Poor readers have difficulty storing and processing information in STM. Since the attention (mechanism) and memory size change as children grow older, this study will investigate how children's reading abilities, which influence the children's perception of text difficulties, change over four years. The psycholinguistic model (so named because psycholinguistic factors are used in the model) in this study takes into account the limitations of STM capacity.

To investigate how STM is related to reading comprehension, test items are analyzed according to three psycholinguistic variables: sentence length, word frequency, and propositional density. According to Baddeley et al. (1975), the phonological loop in immediate memory performance is directly influenced by the spoken length of memory items. In this sense, using length of word for determining sentence difficulty is related to the efficiency of STM or working memory. Especially when considering that beginning readers undergo a decoding stage and that the children in this study are beginning readers in 1988, the first year of the data collection, the phonological loop in working memory is assumed to be involved in children's early stage of oral reading. Also, familiar words do not take much memory space because of the effect of automaticity. Propositions, psychological representations of meaning, are composed of a predicator and arguments (Kintsch, 1974). For example, the sentence, "John runs fast." consists of two propositions: [run, John] and [fast, run]. In addition to this, the more propositions in a sentence, the more STM space they require since the number of propositions is comparable to the number of conceptual meaning units or memory chunks. In this sense, the use of these three variables is directly related to the capacity of STM or working memory. However, in order to avoid a probable multicollinearity between sentence length and number of propositions, the density of propositions, which is obtained by dividing the number of propositions by the number of words in a sentence, will be used.

In addition, studies of early literacy show the importance of home environment and intra-individual characteristics. However, for theoretical consistency, variables such as home cognitive stimulation score and variables that reflect intra-individual characteristics will be used to investigate how children's reading ability and the relative contribution of psycholinguistic variables change over time. Children from enriched home environments and children who have high verbal memory typically demonstrate better reading achievement. This study will investigate the pattern of the children's ability while controlling for intra-individual and home environmental factors. In addition, this study will examine the relative impact of each cluster of variables on children's reading development, while controlling for other contextual characteristics.

The methodology employed in this study, HGLM, provides a vehicle to evaluate my research questions. In the HGLM to be used in this study, item responses (which have linguistic characteristics) are nested within testing occasions. Occasions, in turn, are nested within individuals who differ from one another on several characteristics. The following are the specific research questions:

1. Do children's reading abilities change at a constant rate?
    a) Do abilities increase at constant or variable rates over time?
    b) Do changes in reading abilities differ across individuals?

2. How does the importance of each linguistic/psycholinguistic variable change as children grow older?

   Is the rate of change for each linguistic variable constant or variable?

3. How do individual children's characteristics such as verbal memory interact with text characteristics, such as sentence length, vocabulary frequency, and propositional density?

   a) Does the effect of sentence length on reading comprehension depend on children's verbal memory?

   b) Does the effect of vocabulary frequency on reading comprehension depend on children's verbal memory?

   c) Does the effect of the propositional density on reading comprehension depend on children's verbal memory?

4. How do contextual factors influence children's growth in reading?

   To what extent does children's growth in reading depend on

   a) verbal memory?

   b) home environment?

   c) race?

   b) gender?

   d) the initial test month?

CHAPTER 2

LITERATURE REVIEW

This study draws on three relevant bodies of literature related to the use of the Peabody Reading Comprehension Test. The first is the information processing view of cognitive processes, including reading. The second is the long-standing empirical tradition of estimating the readability of text by examining its linguistic characteristics. The third is a developmental stage-wise view of reading, one that suggests that the cognitive demands of reading change as the task increases in complexity.

Theoretical Perspectives

Miller (1993) argued that information processing is not a single theory, but rather a framework which characterizes a large number of research programs. The flow of information begins with an input, or stimulus. It ends with an output, which could be a bit of information stored in long-term memory (LTM) or an observable behavior such as a speech act or a decision of choosing one answer over another. Since mental operations occur in short-term memory (STM) during the real time between input and output, the consideration of STM (or working memory) is useful for this study.

William James (1890) proposed that the essence of attention is focalization, concentration, and consciousness. Attention requires withdrawal from some things in order to deal effectively with others. Because of the limited capacity for attending to stimuli (Broadbent, 1958; Treisman, 1960; Posner, 1982), performance may break down if the attentional demands of the task exceed the performer's capacity (Anderson, 1982). However, as practice increases, performance becomes more automatic, requiring less attention (Laberge & Samuel, 1974) and less STM or working memory space. Chunking, which can be regarded as organizing stimuli into a meaningful unit, is also related to automatization in the sense that the perceptual system rapidly parses the stimulus, forming a hierarchical structure of instantiated chunks (VanLehn, 1989).

Miller (1956), observing that STM has a limited capacity, posited his now famous 7 ± 2 rule, specifying that STM can only deal with about seven chunks of information concurrently. According to Miller, although the size of a chunk might differ among individuals, the number of chunks remains the same. However, his conclusion is based on research with adults; children's memory chunks are smaller and change both quantitatively and qualitatively as they develop. Two general sources of changes in processing are the acquisition of particular cognitive skills and increases in the capacity or rate of processing (Miller, 1993).

Baddeley and Hitch (1974) presented a working memory model in which there are three components in working memory: a central executive component, a phonological loop, and a visuo-spatial sketch pad. The central executive component regulates information flow within working memory, retrieves information from other memory systems such as LTM, and processes and stores information. However, the processing resources used by the central executive are limited in capacity. The efficiency with which the central executive fulfills a particular function depends upon whether other constraints are placed on it (Gathercole & Baddeley, 1993).

The central executive is supplemented by two components or slave systems--the phonological loop and visuo-spatial sketch pad. The phonological loop maintains verbally coded information, whereas the visuo-spatial sketch pad is involved in the short-term processing and maintenance of material which has a spatial component. These two systems as well as LTM size undergo changes as children grow older. A study by Gathercole et al. (1991) showed that the phonological loop is related to verbal memory and vocabulary knowledge. A study by Scarborough (1998) showed that kindergartners' verbal memory score is more strongly related to their future reading achievement than digit span, word span, and pseudo-word repetition measures. This study will investigate how much the verbal memory obtained around the age of four influences children's reading abilities over three points in time.

Changes in reading ability may come about through certain kinds of experiences. Some experiences are stored as schemas or scripts in the LTM, which can be brought into the working memory when needed. For example, schema theory explains that text comprehension varies directly with experiential background--that readers can easily understand text when it matches their experience (Anderson & Pearson, 1984). Experiences include encountering conflict between different predictions, becoming more familiar with the task materials, trying out a strategy that works, and acquiring more knowledge about the physical and social world (Miller, 1993). These experiences lead to new rules or strategies, which in turn lead to better memory, representation, and problem-solving. In this sense, experience is one major factor inducing cognitive development. However, the social environmental experience is not the initial or central interest of information processing theory (Gardner, 1987) although numerous studies have shown the importance of home environment on children's cognitive development. In the NLSY data set, *home score*, which is the combined score of *cognitive stimulation* score and *emotional* *support score*, exists. However, for theoretical consistency, I used home cognitive stimulation score to investigate the effect of home environment on the children's reading ability over time. In addition, through this study, I will investigate whether individual differences exist after controlling for a changing environmental factor and individual differences.

Need for Readability Study

Attainment of literacy in reading is directly related to academic, economic, societal, political, and personal life and values (Harris, 1990). As far back as 1935, Gates described reading as the most important and the most troublesome subject in primary schools. Since mastering reading is essential to learning almost every other school subject, failure in the primary school is directly related to deficiencies in reading. Along the same line, Ogle, Absalam, and Rogers (1991) reported that students who have difficulty in reading are more likely to experience unemployment upon leaving school. Reading is a vital developmental task that should be mastered. Recently, national attention has been drawn to reading, or more precisely reading disabilities; a report issued by the National Research Council (Snow & Burns, 1998) showed the devastating consequences of a reading disability. In most cases, unsatisfactory achievement in reading has a handicapping effect on an individual's life.

Because of the importance of reading, for decades researchers have tried to find various ways to improve students' reading ability. Numerous individuals and commissions have offered their analyses and recommendations to improve reading. Texts were the central aspect in these reports and emphasis on quantifiable standards brought renewed interests in readability studies (Bruce & Rubin, 1988). Research studies (Hahn, 1987) also showed that if texts are too difficult, children exhibit behavioral problems during class by being less attentive. Carver (1994) also implied that easy text books, which are characterized by the existence of less than 1 percent of unknown words, are not appropriate for enhancing children's vocabulary. Thus, an optimal level of text difficulty is needed to induce children's learning. Developmentally appropriate texts are neither so easy that they offer no challenge to children, nor so difficult that children feel frustrated. The prediction of text readability has been championed as a tool to enhance or maximize students' learning because it affords the selection of developmentally appropriate texts. However, no studies have been conducted to investigate the importance of text characteristics over time, especially with the same students across several years.

History of Measuring Text Difficulty

According to Klare (1985), readability concerns itself with qualities of writing which are related to reader comprehension. Readability formulas refer to a predictive device (Klare, 1963) intended to provide quantitative and objective estimates of reading difficulty (Klare, 1985). Readability formulas have been used as an indicator of comprehension difficulty of reading materials (Carver, 1977-78).

Readability has been studied in two traditions, prediction and production. In the prediction tradition, readability of a text has been investigated to predict how readable a piece of writing is likely to be for the intended reader or to predict the grade level of the written materials. In the production tradition, readability of a text has been manipulated experimentally to produce readable texts for readers in a target population. Prediction research has been done by applying psychometric theory, where the validity and reliability have been high compared to production research studies. The prediction research studies can be generalizable because a large sample size of the criterion variable is used. However, production research studies, which are done in the psycholinguistic tradition, have comparatively low reliability, which influences their replicability and validity. As production research studies are implemented experimentally, they can be used to test causal inferences regarding the effects of particular texts. Even so, results of text experiments are often questioned on grounds of generalizability to a population of passages because of the small number of sample passages in a given study (Klare, 1984).

According to Klare's (1963, 1974-5, 1984) historical accounts of readability measurement, the development of readability formulas goes back to the early 1920s. H. D. Kitson (1921) can be considered as its pioneer. He used the number of syllables in a word and the number of words in a sentence as indices of the relative difficulty of newspapers and magazines. Since then, numerous readability formulas using linear regression have sprung up. Among them, Lively and Pressesy's formula (1923) used a word frequency index based on Thorndike's Teacher's Word Book to estimate vocabulary difficulty. Lodge's (1939) formula used semantic and syntactic factors, which are still the most widely used variables.

Flesch's (1943) formula was designed for adult materials. According to Flesch, formulas then existing were not fit for adult materials because of their emphasis on vocabulary frequency at the expense of other factors. Flesch's formula put emphasis on abstract words. Using magazine articles as criterion variables, he found that counting abstract words and affix morphemes[1], as a means of measuring abstractness, was closely related to the magazine levels. However, the tediousness of counting affixes as a means of measuring abstractness and the often misleading methods of counting personal references led to the development of two formulas. One of them is the most popular, Flesch's Reading Easy Formula. This formula used the number of syllables in a word and the number of words in a sentence as indices of syntactic difficulty of a systematically selected 100 word sample of materials. (Klare, 1963/1984). The formula correlated 0.70 with the McCall-Crabbs criterion. The other formula is Flesch's Human Interest Formula, which used personal words per 100 words and personal sentences per 100 sentences. Personal words means using personal names instead of using proper noun. For example, "**Mike** said that...." Personal sentences are those sentences aimed directly at readers. For example, "**You** should do...." This formula correlated 0.43 with McCall-Crabbs criterion. To supplement some deficiencies found in Flesch's original formula, Dale and Chall (1948) used familiar words to determine semantic difficulty using Dale's list of 3,000 words and sentence length (in words) in their formula. Dale-Chall formula scores correlated 0.70 with McCall-Crabbs criterion scores (which is based on multiple choice, and has been widely used as a measure of comprehension). Dale-Chall's formula is highly predictive of text difficulties.

Gray and Leary's (1935) work was also salient because of its comprehensiveness and the methods of conducting factor analysis for building a formula. This formula is also

[1] Affixes are the additions to stems, roots, and words to modify the meaning of words. For example, **im-** in **impossible** is used as a prefix and **-ness** in **goodness** is used as a suffix.

intended for adults. Gray and Leary (1935) employed survey methods to isolate factors contributing to readability. Existing work and surveys of experts' opinions and reactions of library patrons yielded 289 factors. These are grouped into four major categories such as content, style of expression and presentation, format, and general features of organization. To understand adult reading ability, they developed the Adult Reading Test and found that 44 factors out of the 82 style factors were significantly related to reading score. Due to high correlations among these 44 factors, five of these factors -- number of personal pronouns, number of words per sentence, number of prepositional phrases, and number of different hard words-- were singled out to be used in the readability formula.

Most formula developers used children's material in the developmental process, which raised validity issues (Klare, 1975). However, Flesch's, Gray and Leary's, and Dale-Chall's formulas were intended for adult materials. Some formulas also yielded grade-level scales. For example, the Fox Index developed by Gunning (1952), the Degrees of Reading Power (which can be rescaled into Grade equivalent units), and Stenner's lexile scale all yielded grade level estimates of difficulty. Some of these programs and the research underlying them will be discussed in a later section.

There are several authors who measured text difficulties without relying on readability formulas: clinical approaches, tests, and cloze procedures[2]. The clinical or individual approach was also frequently used as a means of measuring readability (Klare,1963). For example, Dewey (1931) interviewed children to understand the nature and limitations of comprehension in reading history. However, due to subjective judgment

[2] Cloze procedure is the deletion of words in a text at stated intervals, in which readers are asked to fill in words correctly (Zakaluk, & Samules, 1988).

that is prone to errors, the clinical approach is often used in conjunction with the readability formula. Tests are also used for measuring text difficulties. However, constructing and administering a test is a difficult and time-consuming process compared to predicting readability. Taylor (1953) developed the cloze procedure, which requires students to fill in blanks of a text that appear after every few words, usually every five words. Klare (1963) criticized the cloze procedure saying that it is not a formula. However, it is a quick and easy testing technique that may be used for developing criteria in the construction and validation of readability formulas. Unlike traditional readability formulas which do not require testing of human subjects to provide readability scores for passages, the cloze procedure does take into account the reader factor (Klare, 1984).

However, Carver (1977-78) criticized the cloze test because the cloze difficulty estimate depends on the ability level of the particular group to whom the test was administered as well as the difficulty level of the material. Even when an ability adjustment for cloze was developed, it was still an impractical method in many situations because it was always necessary to have a norm group before a language difficulty estimate was obtained (Carver, 1977-78).

The most comprehensive exploration of variables was completed by Bormuth (1966). Using correlation and regression, Bormuth (1966) explored more than 100 structural variables. Among them, more than 60 variables were significant in predicting comprehension difficulty of a criterion variable which was measured by the cloze test. According to Pearson (1969, 1974-75), Bormuth's contribution in the area of readability was significant in that he was able to estimate readability using multiple regression at the level of word (R=0.51), the independent clause (R=0.67), the sentence (R=0.68), and the passage (R=0.93), whereas traditional formulas cannot be reliably applicable to below passage level. In addition to this, Bormuth's exploration of the parts of speech ratio significantly predicted text difficulty. For example, he found highly explanatory linguistic ratios, such as pronoun/conjunction (r =0.81), interjection/pronoun (r =0.62), and verb/conjunction (r =0.73). He also used quadratic terms in his regression model and showed the existence of a nonlinear relationship between outcome variables and a predictor. In his study, Bormuth also applied Yngve's (1960) word depth analysis as a means of measuring sentence complexity. According to Yngve, the notion of word depth comes from mechanical translation of language by electronic computers. Embedded sentences, such as "the cat that the dog chased was gray," require more memory because the machine has to store information from the beginning of the sentence (the cat) up to the end of the sentence (was gray).

However, Bormuth's use of many variables was not based on any consistent theoretical perspective. Bormuth's major concern seemed to be in the explanatory power of variables such as sentences length, parts of speech ratio, and depth of words. Pearson's (1969) summary on the variables found in 31 readability formulas, which was mentioned in Klare (1963), showed that word frequency (18), sentence length measure (17), number of syllables (9), sentence complexity (9), and conceptual measure (10) were widely used.

As was seen in many earlier readability formulas, text difficulties have been measured by semantic and syntactic factors. Among semantic factors, vocabulary difficulty was one of the most significant predictors of text difficulties (Dale, 1965; Davis, 1968; Chall, 1983). As a measure of syntactic difficulties, sentence length or word length has been frequently used. However, short sentences do not necessarily make a text easy to comprehend (Chall, 1958; Klare, 1963; Kintsch, 1979; Pearson, 1969). Besides this, using factors other than semantic and syntactic was not successful in predicting text difficulty. A recent study by Stenner (1997) using the PIAT reading comprehension test showed that the combination of sentence length (the log of mean sentence length) and word frequency (the mean of the log word frequencies) explained 85 percent of the variance in the PIAT item rank-order difficulty. However, Stenner's study did not incorporate the effect of word order or syntax which has been shown to operate somewhat independently of sentence length (e.g., Pearson, 1974-5). Notice also that there are some sentences in which sentence length cannot be a genuine explanatory factor: If we were to scramble the order of words in a sentence, it could be difficult or even incomprehensible even though sentence length had not changed at all.

Readability formulas have not had strong theoretical perspectives (Kintsch, 1979), and formulas have been based on apparent, or surface level, text characteristics. For example, Bormuth's (1964/66) exploration of more than 60 variables which contributed to the variance of the criterion variable, using the cloze test, was not based on a consistent reading theory, although some of these variables seemed quite reasonable and plausible.

Before Kintsch's readability formula (1979), which incorporated some aspects of the psychological processes of the reader, most readability formulas confined themselves to measuring observable text characteristics. Most traditional readability formulas have not directly taken the reader's ability into account. According to Baker, Atwood, and Duffy (1988), the traditional readability model regards the process of reading as a passive activity, in which the reader decodes the text to obtain meaning. Therefore, reading can be defined in terms of the skills necessary to decode words and sentences. Because reading is viewed as decoding words and sentences, the difficulty of the text is indexed in terms of word (lexical features) and sentence characteristics.

If literacy is determined by the reader's ability as well as the difficulty of the text (Bormuth, 1966), then the earlier formulas are problematic because they do not take into account the reader factor (Kintsch & Vipond, 1979). According to Bruce and Rubin (1988), readability formulas have limitations because formulas do not measure all the factors that influence the comprehensibility of a text. Since existing formulas have measured only one aspect of writing, the difficulty of style, they have not touched content, organization, word order, format, or imagery of writing; nor have they embraced reader factors such as purpose, maturity, or intelligence (Klare, 1963). A good readability score does not mean that the piece of writing was written well. Formulas have not taken into account other elements such as content, or other aspects of style, such as mood. In addition to this, the traditional readability grade level index found in traditional readability formulas produced different results (Bruce & Rubin, 1988). A grade level score for an individual based on a typical reading test means that he/she reads as well as some normative group. Along with this, in the traditional readability study, reading is viewed as a general process independent of domain knowledge. The typical formulas are applied regardless of the nature of tasks, subject, and expertise of reader (Baker, Atwood, Duffy, 1988).

However, Kintsch's readability approach is different. Kintsch (1979) regarded readability "not as immutable property of text, but as the result of a reader-text interaction." Unlike traditional readability formulas, Kintsch's model is based on information processing theory. His model came out of empirical observations such as recall or text processing time. In his model there are two given conditions: the reader, who usually has a goal schema to understand the text or at least to find out what is new in it and the text, which is represented as propositions. Examining text as a semantic representation, Kintsch codes the text into a set of propositions or conceptual structures that represent the meaning of the text. Kintsch wanted to identify the process that occurs between input propositions (lowest level) and readers' goal schema (highest level). The lowest level of propositions is needed to predict a part and the level of the input propositions that people recall. The input propositions construct a coherent network, identifying places where inferences are required to obtain coherence. To predict the summaries that people make of a text, the hierarchical macrostructure is also needed. In this model, information flows both bottom-up and top-down. According to Kintsch, to connect new information with old, readers need to search for old information, which is called reinstatement search. If readers have to make a large number of reinstatement searches and a large number of inferences, then reading will be difficult. Based on this model, Kintsch's readability formula puts such variables as number of reinstatement searches made by the model in processing the paragraph, the average word frequency, propositional density, the number of inferences, the number of processing cycles, and the number of different arguments in the proposition list. The first two variables--reinstatement searches and word frequency-- explained most of the variance, but all six variables together explained 97 percent of the variance of the outcome variable, recalling the text.

The role of propositions was also investigated by Pearson's experimental study of the reading process with above average 3rd and 4th grade readers. According to Pearson

(1969/ 1974-75), readers do not process a text analytically as was indicated by transformational grammarians. Transformational grammarians think that if the sentence we read or hear is close to the deep structure (the meaning), then less transformation is applied, which facilitates comprehension. Pearson's study also did not support the idea of traditional readability studies which show the length of the sentence as a significant index of readability of texts. As was indicated by Klare (1963/1984) and Kintsch (1979), Pearson's study also implies that reducing the length of a sentence does not necessarily facilitate children's recall of text. Instead, children try to make a coherent whole when they process text, which is more consistent with propositional analysis.

Studies conducted in the psychometric tradition have incorporated both reader's ability and characteristics of texts. Carver's (1977) and Stenner's studies (1997) took into account both the reader's ability and text difficulty. Carver (1977-78) maintained that the prediction of reading comprehension is made by the ability level of the reader and the characteristics of text. In traditional readability studies, ability levels were often scaled using standardized tests and these measures initially were not scaled with respect to the difficulty of the text (Carver, 1978-77). In Carver's (1977-78) National Reading Standards, each grade ability score on the test (Ga) had been calibrated to reflect a 0.50 probability that an individual can read and understand, or comprehend the passages at the same grade of difficulty (Gd) according to the Rauding scale. The Rauding scale measured the grade difficulty of reading and understanding. A grade 5 ability means that the average accuracy is likely to be 75 percent of grade 5 materials. A choice of a 75 percent target comprehension rate is obtained through empirical evidence (Square, Huitt, and Segars, 1983; Crawford et al., 1975). The theoretical assumption of comprehension in using the Rauding scale is that the rate of reading is constant and the accuracy of comprehension during reading can be predicted from a measure of material difficulty and individual ability. However, the Rauding theory was criticized because it is very mechanical, serial, and not comprehensive. In this sense "the theoretical assumption does not support every day reading phenomena such as skimming and studying (Pearson, 1977-78).

Stenner's (1997) study on the Lexile framework (reading comprehension scale) also took into account both the reader's ability and text difficulty. In order to obtain generalizability, that is, the scale of a single object being independent of conditions, scores obtained from different test administration should be tied to a common zero (anchor). To obtain general objectivity, theoretical logit difficulties obtained were transformed to scales that could be compared to each other without ambiguity. Measurements for all persons and all texts are reportable in a Lexile framework.

Some studies which investigated developmental aspects of children's reading used grade appropriate assessments using a cross-sectional design. These studies employed linear models using GE (grade equivalent) scores that were extrapolated beyond the grade that were actually assessed (Klare, 1984; Chall, 1970)[3]. However, no studies have been done to measure both the rate (acceleration/deceleration) of readers' ability and the relative importance of each text characteristics over time using reading materials that can accommodate a wide range of readers.

Gray and Leary's (1935) and Bormuth's (1964) studies provided evidence that

[3] Extrapolation beyond the grade level that was used in the criterion measure is not a valid assessment.

linguistic variables do not predict comprehension difficulty equally well for subjects with different levels of achievement. Besides, Draper et al.'s (1971) study and Chall et al.'s (1990) study indicated that vocabulary explains text difficulty better at more advanced than at early stages of reading development. This present study investigated how the importance of the psycholinguistic/linguistic characteristics of text changes across years. In this study, in addition to the most popular variables--frequency of vocabulary and length of sentence--propositional density was used to investigate the significance that propositions play in the readability formula at each time point.

Reading Development

A Developmental Perspective

Chall (1983) categorized six developmental stages, from stage 0 to stage 5, which characterize prototypical reading development. According to Chall, stage 0 is a prereading stage covering birth to age 6. At this stage a child gains some insight into the nature of words before going to school. Stage 1 is an initial decoding stage covering grades 1-2 (6-7 years old). A child associates arbitrary letters that they learn with the corresponding parts of spoken words. Stage 2 covers grades 2-3 (7-8 years old). At this stage, the child reads not for gaining new information, but for confirming what is already known. Children pay attention to the printed words, usually the most common and high frequency words. Stage 3 reading is also characterized by the growing importance of word meanings and of prior knowledge. This stage is composed of two phases: Phase l of stage 3 covers grades 4-6 (9-11 years old) and children develop the ability to read beyond an egocentric purpose, reading texts that convey conventional knowledge of the world. Phase 2 of stage 3 covers grades 7-8 (12-14 years old): This stage brings readers close to the ability to read on a general adult level. Stage 4 reading is characterized by a child's capacity to adopt multiple viewpoints. This stage covers high school grades (14-18 years old). Stage 4 is mostly acquired through formal education. Stage 5 covers college level and is characterized by construction and reconstruction of a world view. Since the NLSY children in this study undergo three reading developmental stages, starting from stage 1 to stage 3, it provides a great opportunity to investigate beginning reader's reading development although it must be conceded that the PIAT does not lend itself to even a weak test of the validity of Chall's stage theory.

Contextual Variables

The 1994 NAEP (National Assessment of Educational Progress) reading assessment shows that contextual influences, such as school and home environment, affect children's reading proficiency. However, it is assumed that the effect of these contextual variables may differ as a function of the developmental level of children. Luster and Dubow's (1992) study of environmental factors on children's verbal intelligence shows that the effect of environment changes depending upon the children's developmental level. Evidence from an adoption study by Plomin and Daniels (1987) also indicates that the effect of shared home environment is reduced as children grow older, while the effect of the non-shared environment, such as schooling effects, becomes greater. In this sense, a developmental study is needed to investigate differential effects of contextual factors. To understand the effect of changing home environment on children's reading abilities, home cognitive stimulation score will be used. Because of access to the larger NLSY database, the effect of other intra-individual factors such as gender, race, verbal memory, and testing time, will be investigated.

Operationalization of the Factors in the Present Study

Building on information-processing theory, this study will investigate both factors that are internal to the text, such as linguistic and psycholinguistic variables, and factors that are external to the text, such as individual differences among readers. Children's internal characteristics, such as verbal memory, are used in order to investigate the pattern of reading development, while controlling for their effect on the growth of children's reading ability.

Understanding children's reading development is related, at least indirectly, to the item development process underlying the PIAT. A better understanding of children's reading development would be one of the essentials for selecting and constructing the crucial subtest and its items. Norm referenced tests could benefit from a better knowledge of the qualitative changes in reading (Chall, 1983). Although the PIAT reading comprehension test has certain limitations, especially because the text of each item consists of one single sentence, it will also show various characteristics that children face in understanding texts at different time points.

Thus far I have discussed information processing theories, linguistic and psycholinguistic correlates of text difficulty, particularly as they are related to readability formulas and matters of reading development, as they are reflected in individual differences among children. The statistical models used in the current study permit me to investigate each of these potentially important sources of variation. For example, the level-1 model represents the nesting of test items within each occasion (3 time points across 4 years) and affords the evaluation of linguistic/psycholinguistic variables; the level-2 model represents the nesting of occasions within a child which measures pattern of development and changing environmental effect on a child's reading development; and the level-3 model represents the intra-individual characteristics. By building the model from a lower to a higher level, I can investigate how the importance of each variable changes across occasions; how individual children's reading ability changes due to the changing environmental characteristics; and how time-invariant individual characteristics influence the development of an individual child's reading ability.

CHAPTER 3

METHODOLOGY

Subjects

The subjects for this study are 477 children from the National Longitudinal Survey of Youth (NLSY) data set, chosen based on age and scores on the PIAT Reading Recognition Test. Children's ages ranged from 6.0 years to 6.11 years in 1988. There were 220 boys and 257 girls, among them, 89 Hispanic children, 153 Black children, and 235 non-Black- non-Hispanic (White) children. The children's responses to reading comprehension items were observed over three time points, approximately every two years, 1988, 1990, and 1992. Those who scored over 15 on the PIAT Reading Recognition Test were given the Reading comprehension test. These are the children in the sample for this study . According to Chall's (1983) developmental scheme, which divides children's reading development into six stages ranging from 0 to 5, the NLSY children in 1988 would be roughly categorized into stage 1, and can thus be defined as beginning readers.

Children who took the PIAT Reading Comprehension tests were the offspring of individuals selected for the National Longitudinal Survey of Youth (NLSY '79) project. The NLSY mothers have been interviewed annually since 1979, when they were 14 to 21 years of age. The NLSY '79 child sample, when weighted, represents a cross-section of children born to a nationally representative sample of women who were between the ages of 29 and 36 on January 1, 1994 (NLSY, 1997). It is estimated that the children in the sample typify approximately the first 70 to 75 percent of children born to the contemporary cohort of American women (NLSY, 1997). The original NLSY '79 sample included 6238 women in 1979, 456 of whom were in the military at that time. However, none of the subjects in this study were from these mothers because most of them were dropped before my data collection. In addition, children born to the economically disadvantaged White women were not available because of financial constraints of the NLSY project. Every two years from 1986 to 1994, a series of assessments were administered to the children of NLSY mothers as a means of measuring the children's cognitive ability. Children of Hispanic, Black, and non-Hispanic and non-Black (White) ethnic groups of both sexes were investigated for this study. Data up to 1992 were gathered primarily in person using paper and pencil assessment techniques. However, information about children's item responses was not available in the 1986 data. Also, due to large attrition, the 1994 data were not included in this study. Thus, the result can only be generalized to the population with the above characteristics.

Outcome Measure

General Characteristics of the PIAT Reading Comprehension Test

The Reading Comprehension test in this study is one of five subtests from the Peabody Individual Achievement Test Battery: Mathematics, Reading Recognition, Reading Comprehension, Spelling, and General Information. However, the NLSY data has information only about three subtests: Mathematics, Reading Recognition, and Reading Comprehension. The PIAT Reading Comprehension test was designed for children in kindergarten through grade 12. It was originally intended for children scoring age 5 years and over on Peabody Picture Vocabulary Test (PPVT) and at least 19 on the Reading Recognition assessment.

Interviewers in the NLSY study administered the PIAT Reading Comprehension tests to children whose Reading Recognition score was over 15. Scores were calculated by deducting the number of incorrect responses from the ceiling item number--the highest numbered (in a sequences from easy to hard) item that the child missed. Children who scored less than 19 on the Reading Recognition test were assigned their Reading Recognition score as their Reading Comprehension test score. Total raw scores ranged from 0 to 84. The PIAT Comprehension test item number ranges were item number 19 to item number 84 (total 66 items).

The PIAT Reading Comprehension sub-test measures children's ability to derive meaning from sentences that are read silently (Dunn & Markwardt, 1970). Item construction was based on the assumptions that "reading is the facility to derive meaning from printed words" (Dunn & Markwardt, 1970) and that the effective reader can retain the meaning after exposure to the illustrations in the absence of the passage. Thus, the PIAT Reading Comprehension Test is highly memory dependent.

The individually administered test is composed of 66 one-sentence items of increasing difficulty. According to Dunn and Markwardt (1970), difficulty is based on sentence complexity, vocabulary, and sentence length. The child silently reads a sentence displayed on a separate page, the interviewer shows the child four pictures on the other side of the page, and the child is asked to select the correct picture. The PIAT Reading Comprehension test is a recall type of reading comprehension assessment because the children are asked to select, without reading the text again, the one picture that best depicts the sentence. In other words, the PIAT Reading Comprehension test depends heavily on short term memory and attention. It is a combination of a time and power test

(Nunnally, 1978) in that children are encouraged to respond to each item within 30-40 seconds, although Dunn and Markwardt intended this to be a power test.

The PIAT Reading Comprehension has no written directions for the children to respond to each item. In this aspect, the PIAT reading comprehension test eliminates some problems related to validity that might arise from the gap between text understanding and question understanding, as found in other types of reading comprehension tests. Due to misinterpretation of directions or questions in some tests, children may not respond to questions correctly although they understand the body of the text.

The PIAT Comprehension test is an adaptive test. Complete responses to all items are seldom, if ever, collected. Items are arranged in ascending order of difficulty with the easiest questions being comparable to kindergarten or first-grade level. None of the children attempt all of the items. Instead, interviewers test children with the items in the children's critical range by constructing a basal level and a ceiling for each child. A basal level is derived from a series of correct responses, and a ceiling is determined from a series of continuous errors. The basal level is determined by finding the highest cluster of five consecutive items answered correctly. The lowest numbered item in that cluster is designated as the basal item. Most coders for this NLSY data actually coded the highest item number in a set of five consecutive correct items as a basal item. However, this coding mistake did not make any difference in imputing missing values below basal item number. The ceiling is obtained by continuing to present increasingly challenging items, until the subject had made a total of five consecutive errors. The last item missed in the set of five is regarded as the ceiling item. In contrast to the errors made by coders for basal items, most coders applied the procedures for determining ceiling items appropriately. This process is illustrated in Figure 1, where the basal range is from the item 22 to the item 26, and the basal item number is question 22. Ceiling range is from item 31 to item 35 and the ceiling item number is 35.

	Item#	Score	Imputation
	19	1	Imputation
	20	1	Imputation
	21	1	Imputation
*Basal item#	22	**1**	
	23	**1**	
	24	**1**	
	25	**1**	
	26	**1**	
	27	**0**	
	28	**1**	
	29	**0**	
	30	**1**	
	31	**0**	
	32	**0**	
	33	**0**	
	34	**0**	
*Ceiling item#	35	**0**	
	36		Imputation
	37		Imputation
	38		Imputation
	39		Imputation
	:		Imputation
	84		Imputation

Figure 1. Information about forming ceiling and basal items where score = 1 is correct and score = 0 is incorrect.

Information about the basal and ceiling items is available with the NLSY data (information about basal item number is not available in 1988). However, some mis-coding also occurred on the information about basal and ceiling item number. Partly because of the PIAT interviewers' mis-coding, information on ceiling number and basal item number is not always correct. Subsequently, I corrected them for the purpose of imputation. For this study, all the raw reading comprehension item responses were checked one by one to establish ceiling and basal levels for the imputation. If there was no clear-cut information on forming basal and ceiling, the item responses were imputed as missing. However, while recoding this, I found out that some interviewers did not assess children on enough reading comprehension items, and some interviewers gave more opportunities to respond than the procedure calls for. Especially in 1988, interviewers did not give enough opportunities to form a ceiling partly because they could not form basal levels in many cases.

Because of many missing item responses outside of actual item responses, the raw data information was consulted in order to impute scores. Imputations on the items below the basal question (the lowest numbered item in the lowest set of five consecutively answered correct response) were made by assuming that children would answer all lower level items correctly (imputed as 1). Imputation on the items beyond the ceiling item number was accomplished by regarding these to be wrong (imputed as 0). Since the PIAT test is a multiple choice test with four options, if children are given an opportunity to respond, the probability of children's making a correct response by blind-guessing is 0.25. To solve this problem of unequal opportunity, responses to the untried items beyond the top-most difficult item were assigned by randomly generating the real numbers between 0 and 1. If the randomly generated number was greater than or equal to 0.75, the item was imputed as correct (1); otherwise, it was imputed as incorrect (0).

Validity and Reliability of the PIAT Reading Comprehension

The reading comprehension subtest of the PIAT is generally considered to be a highly reliable and valid achievement test, and has been extensively used for research purposes (NLSY, 1992). Because of the format and the high probability that any given child will not complete the entire test, test-retest reliability is the only viable index available to evaluate consistency. According to Dunn and Markwardt (1970), the median test-retest reliability was 0.65 (ranges from r =0.61 to 0.78) and standard errors of measurement for raw scores on selected grade levels ranged from 2.48 (grade 1) to 7.39 (grade 8), which implies that the PIAT is not so reliable for measuring older children's reading abilities.

Dunn and Markwardt (1970) defined reading as a functional ability, the facility to derive meaning from printed words. The reading comprehension test construction was not based simply on finding the meaning of individual words, but on the ability to comprehend passages in context. Although the passages are composed of single sentences of varying length and difficulty, they have content validity, covering kindergarten to grade 12 reading levels. Bormuth's study (1966) also validated the efficacy of assessing sentence-level reading comprehension using multiple correlation with other predictors (R=0.68). Item discrimination and difficulty indices were used for the PIAT. For each item, a curve was drawn showing the percentage of children passing at each successive grade level. Items were retained that showed the sharpest curves, and were placed at the grade level where approximately 50 percent of the subjects passed. Internal consistency was built in by selecting items that correlated most highly with the total score.

Concurrent validity was assessed by examining the correlation between the Peabody Picture Vocabulary Test and the PIAT Reading Comprehension Test. The correlation coefficients ranged from 0.42 to 0.70 across different grade levels. This version was normed in the late 1960s and renormed in 1990. Norms, however, are not a major consideration in this study because raw score growth patterns rather than normed scores are the primary data of interest.

Model and the Predictor Variables

In this study, to understand the nature of growth in reading comprehension, a three-level hierarchical generalized linear model (HGLM) (Bryk, Raudenbush, & Condon, 1996) was used. Item responses (level-1) were considered as being nested within testing occasions (level-2) and testing occasions as being nested within individuals (level-3). Since children took the same test on three occasions, each item was nested within each time point (occasions). In addition, the time (in month) that children took the test varied and sometimes occasions (frequency of taking the test) also varied, so it can be considered that time points were nested within individuals. In this study, ability and text characteristics were put into the model. However, here the scores on children's abilities were not obtained directly, but abilities were regarded as an intercept in the HGLM model, when all the text characteristics and other contextual effects were controlled for. By building the model in this way, this study investigated how the level of intra-individual characteristics influence the importance of each item variable over time. In addition, individual reading ability was observed while controlling for changing home environmental factors. Also, children's reading abilities and the importance of each psycholinguistic variable at each time point were observed while controlling for the time-invariant individual characteristics at the level-3 model.

The HGLM can assess the probability of binomial data, which the hierarchical linear model (HLM) cannot estimate. In addition to this, the hierarchical model affords investigation into the contextual effects that influence individual development (Bryk & Raudenbush, 1992). Although the NLSY data contains some missing values, the HGLM can deal effectively with the problem of missing values in the level-1 model. In the case of level-2 and level-3 models, the HGLM program does not allow missing data. For cases in which there were missing value for level-2 or level-3 variables, scores were imputed for each subject based on existing information. In a later section, this procedure will be discussed in detail.

The level-1 model examines item characteristics, and seeks to explain performance by references to the linguistic features of the items. The level-2 model estimates the patterns of growth by examining performance across occasions, in other words, by putting time factors into the model. The level-3 model incorporates the intra-individual characteristics, such as gender, race, and verbal memory. The goal of this analysis is to find the probability, p_{ijk}, of a correct response by child k at one particular occasion j on an item i with specified characteristics.

Since the outcome of the PIAT reading comprehension item was binomially distributed (Bernoulli distribution), a transformation of the probability of responding (the log-odds of response) was used. Because of the nature of the distribution of the dichotomous outcome using the logit model, the probability can be estimated more reasonably. If logit is a linear function of other variables, the outcome, p_{ijk}, is a nonlinear, S-shaped function with the probability range between 0 and 1 (Hamilton, 1992; Bryk, Raudenbush, and Condon, 1996).

Level-1 Model: Item Text Characteristics

The level -1 model in HGLM consists of three parts: (a) a sampling model, (b) a link function, and (c) a structural model. The sampling model in level-1 HGLM is as follows:

1) $Y_{ijk} 1 P_{ijk} \sim \mathrm{B}\ (n_{ijk}, P_{ijk})$

It denotes that Y_{ijk} has a binomial distribution with n_{ijk} trials and probability of making correct response, P_{ijk}. Y_{ijk} is 1 if a person k's response on the item i at time point j is correct; Y_{ijk} is 0 if a person k's response on item i at time point j is incorrect. According to the binomial distribution, the expected value and variance of Y_{ijk} are

2) $\mathrm{E}\ (Y_{ijk} 1 P_{ijk}) = n_{ijk} P_{ijk}, \quad Var\ (Y_{ijk} 1 P_{ijk}) = n_{ijk}\ P_{ijk}(1 - P_{ijk}\ )$.

When the $n_{ijk} = 1$, Y_{ijk} takes on values of either zero or unity which is a Bernoulli distribution. Unlike the Hierarchical Linear Model (HLM), the HGLM allows estimation of models both $\eta_{ijk} = 1$ (Bernoulli case) and $\eta_{ijk} > 1$. For the Bernoulli case, the predicted value of the binary outcome, Y_{ijk} is equal to the probability of making a correct response, $P_{ijk} = \mu_{ijk}$. When the level-1 sampling model is binomial, the HGLM uses the logit link function. $\eta_{ijk} = \log\ (P_{ijk}/1 - P_{ijk})$. η_{ijk} is the log of the odds of making a correct response.

While P_{ijk} is constrained to be in the interval (0,1), η_{ijk} can take on any real value. Predicted log-odds can be converted to predicted probabilities by computing $P_{ijk} = 1(1+ exp^{(-\eta ijk.)})$ Thus, whatever the value of $\eta_{ijk,}$ this procedure will produce a P_{ijk} between zero and one.

3) $\eta_{ijk} = P_{0jk} + P_{1jk}(\text{sentence length})_{ijk} + P_{2jk}(\text{vocabulary frequency})_{ijk} +$

$P_{3jk}(\text{propositional density})_{ijk}$

Here,

p_{0jk}: ability of a child k, at time point j, controlling for item level sentence characteristics

P_{1jk}: effect of sentence length of child k at time point j, controlling for other sentence characteristics of item

P_{2jk}: effect of vocabulary frequency of child k at time point j, controlling for other sentence characteristics

P_{3jk}: effect of propositional density of child k at time point j, controlling for other sentence characteristics of item

At level-1, the probability of child k's response to a certain item is the function of item characteristics such as sentence length, vocabulary frequency, and propositional density. These variables were grand-mean centered (the mean of the average of each predictor), so that P_{0jk} is the probability in log-odds that a child answers an average item correctly when all item characteristics are controlled. P_{0jk} can therefore be considered a measure of ability on the log-odds metric.

Three variables were selected because research studies (See chapter 2) show the selection of these variables as appropriate. Vocabulary difficulty and sentence length are the most widely used variables in readability formulas. Stenner's (1997) study of the PIAT reading comprehension test shows that log of the mean sentence length and the mean of the log word frequencies combined explain 85 percent of the variance ($r = 0.92$). As some previous studies (Shankwiler & Crain, 1986; Stenner, 1997) indicated, the correlation between item rank-order difficulty and sentence length was the highest among the linguistic/psycholinguistic variables. The correlation between item rank order difficulty and sentence length was 0.91 ($R^2 = 0.83$).

For this study, I selected sentence length, vocabulary frequency, and propositional density. Because the raw data were not as skewed as when I log transformed, using raw data, I found that the correlation between item rank order difficulty and sentence length was the highest among all the linguistic/ psycholinguistic variables that I used for this study ($r = 0.91$). Sentence length ranged from 5 to 31 words. The average sentence length was 14.04.

Vocabulary difficulties were measured by the Standard Frequency Index (SFI) based on the total corpus used in the Educator's Word Frequency Guide (EWFG) (Zeno, Ivens, Millard, & Duvvuri, 1995). The most frequently used words received high values in the SFI. Instead of using either high or low SFI in a sentence, mean SFI was used for this study. Mean SFI reflects a more contextual effect compared to words with either low or high SFI. Since it is possible to understand a text without knowing the meaning of every single word, I used average word frequency in measuring vocabulary difficulty. In the EWFG Corpus, observed SFI values ranged between 3.5 and 88.3. In the PIAT the range of SFI values was from 20.8 to 88.3. Derivative words which were not found in the EWFG manual were assigned the lowest value of the words from the same origin. Compound words were treated as one word. The mean of average vocabulary difficulty was 63.67 and the average vocabulary difficulty ranged from 49.45 to 72.70. The correlation between item rank order difficulty and SFI average was 0.66 ($R^2 = 0.44$).

Proposition analysis was based on Kintsch (1974). According to Kintsch, propositions represent ideas and language expresses propositions. A proposition contains a predicator and n arguments ($n \geq 1$). Because it was assumed that longer sentences have more propositions, there might exist a high correlation between length of sentence and number of propositions. In fact the correlation between sentence length and number of propositions was 0.92. Therefore, to avoid the problem of multicollinearity, propositional density--obtained by dividing the number of propositions by the number of words in a sentence--was used. The correlation between rank order and propositional density was 0.11. The number of propositions ranged from 1 to 13 and the propositional density ranged from 0.11 to 0.67. Indefinitives, such as *both*, *every*, *some*, *any*, and *everything* were not analyzed as a predicator. For example, the following sentence has two propositions:

The postman must carefully measure **every** package.

(1) (measure, postman, package)

(2) (carefully, 1)

In addition, the genitive cases (e.g., my, your, his) were not analyzed as forming a meaning unit:

Try kicking **your** feet in the brook.

(1) (kick, you, foot)

(2) (try, 1)

(3) (place: in, 1, brook)

Also verbs in idiomatic expressions were analyzed as one unless it had a unique meaning in the sentence:

A windstorm is **making** a ruin of the cottage.

(1) (ruin, windstorm, cottage)

However, since I only counted the number propositions to obtain the propositional density (number of propositions ÷ length of sentence), the method of counting the number of propositions did not unveil distinctive meanings as was seen in the following examples. The following sentences have the same number of propositions, but the meanings were totally different:

(1) A dog bites a man. (bite, dog, man)

(2) A man bites a dog. (bite, man, dog)

Level-2 Model: Age and Cognitive Stimulation Score

In the level-2 model, the level-1 parameters such as the constant (intercept) and variable coefficient (slopes) are modeled as a function of time, which was measured by age in months at three time points. The value of age was centered around the grandmean, so the estimate of the intercept, B_{00k}, B_{10k}, B_{20k}, and B_{30k}, will be approximately the predicted value for a child *k* at time-point two (at about 8.5 years old). At this level, each parameter (coefficient) from level-1 becomes an outcome.

$$P_{0jk} = B_{00k} + B_{01k}(\text{age linear})_{jk} + B_{02k}(\text{age quadratic})_{jk} + B_{03k}(\text{cognitive stimulation})_{jk} + R_{0jk}$$

$$P_{1jk} = B_{10k} + B_{11k}(\text{age linear})_{jk} + B_{12k}(\text{age quadratic})_{jk}$$

$$P_{2jk} = B_{20k} + B_{21k}(\text{age linear})_{jk} + B_{22k}(\text{age quadratic})_{jk}$$

$$P_{3jk} = B_{30k} + B_{31k}(\text{age linear})_{jk} + B_{32k}(\text{age quadratic})_{jk}$$

B_{00k}: expected ability of individual child *k* at age 8.5, controlling for cognitive stimulation score and sentence characteristics of items such as length, vocabulary, and density

B_{01k}: linear growth rate of child *k*'s ability at age 8.5 on a typical item, controlling for cognitive stimulation score

B_{02k}: acceleration effect of child *k*'s ability on a typical item, controlling for cognitive stimulation score

B_{03k}: effect of home cognitive stimulation score for child *k* on a typical item, at age 8.5

B_{10k}: average effect of sentence length for child *k*, at age 8.5, controlling for cognitive stimulation score and the other sentence characteristics of items

B_{11k}: linear effect of age (growth rate) on sentence length slope at age 8.5 for child *k* controlling for all the other variables

B_{12k} : acceleration effect on sentence length slope for child *k*, controlling for all the other variables

B_{20k}: average effect of vocabulary frequency slope for child *k* at age 8.5, controlling for cognitive stimulation score and the other sentence characteristics of items

B_{21k}: linear effect of age on vocabulary frequency slope at age 8.5 for child *k*, controlling for all the other variables

B_{22k}: acceleration effect on the vocabulary frequency slope for child *k*, controlling for all the other variables

B_{30k}: average effect of propositional density slope for child *k* at age 8.5, controlling for cognitive stimulation score and the other sentence characteristics of items

B_{31k}: linear effect of age on the propositional density slope for child *k* at age 8.5, controlling for all the other variables

B_{32k}: acceleration effect on the propositional density slope for child *k*, controlling for all the other variables

Using the level-2 model, this study can measure whether and how the importance of the item characteristics changes across occasions. Earlier readability research suggested the advisability of examining the effect of these variables at different ages. Gray and Leary's (1935) and Bormuth's (1964) studies provided evidence that linguistic variables did not predict comprehension difficulty equally well for subjects with different levels of achievement. Besides, Draper et al. (1971) and Chall (1990) indicated that at the early stage of reading development, knowledge of vocabulary did not explain text difficulty as effectively as it did at the advanced level of development. This study investigated how the importance of the psycholinguistic/linguistic text characteristics changes across years.

In the HGLM level-1 model, the intercept, which represents the individual reading ability, varies randomly. Because it can change across occasions, using the HGLM model, changes in individual ability can be estimated across occasions. By incorporating quadratic terms, this model can estimate the nature of growth more realistically, looking for both linear increments and non-linear spurts and valleys in growth. According to Chall (1970) and Klare (1984), most readability formulas use linear regression equations, which may not capture the true growth pattern. Bormuth (1964/66) suggested the use of nonlinear models in building readability formulas. By including both linear and quadratic terms in the level-2 model, since the observations were made at three time points, it is possible to investigate whether or not reading ability and the effect of linguistic/psycholinguistic variables change linearly or curvilinearly. Also, with this model the rate of growth across adjacent occasions can be assessed.

In addition, since early reading development is influenced by environmental factors such as interaction with parents, I investigated whether any linear or curvilinear trends remain after controlling for the home environment at each occasion. Environmental factors are not the major focus of this research because many existing studies have demonstrated these effects already. Nonetheless, these interactions with the variables of interest are important because they might modulate any interpretations I might wish to make about the target variables. If level-1 represents micro-level text process, level-2 represents the developmental aspect across time. By incorporating changing environmental variables at level-2 and other time-invariant individual variables at level-3, this model assesses patterns of reading development and the perceived difficulty of text characteristics.

There were no missing values in the 1988 data because age was one of the criteria in selecting subjects. However, for the missing value in the level-2 age linear term, imputation was conducted in the following manner: After estimating a regression equation using 1988 data as independent variables, the standard error of regression was used to generate a random error term which was added to the predicted values for missing values in 1990. To complete the imputation for 1992, I used the same method, estimating a regression equation based on 1990 data and added a random error term generated from the standard error of regression. Therefore, imputation for 1990 and 1992 children's age of taking the test was conducted without changing the nature of distributions. The value of age quadratic term was obtained by squaring the age linear term.

Another level-2 variable, home cognitive stimulation score, is a composite of variables, including number of books that the children have, information on the frequency of parents' reading to the children, and number of hours watching TV. To replace the missing values, imputation was conducted by using both *total home scores* and *cognitive stimulation scores* from the other two years as predictors. The home score was the combination of the home stimulation and home emotional support scores. Probably due to some coding errors, there were some cases in which information on home cognitive stimulation was missing, but information on the home score was available. As indicated in Table 1, the correlation between the total home score and home cognitive stimulation score within each year was over 0.85; higher than that of adjacent year's cognitive stimulation scores. Thus, we used the total home score information first for imputing missing values, and then we used the cognitive stimulation score of the other two years as predictors for predicting expected outcomes. I used the home score information first for imputing missing values. Then, I used the other two years as predictors for predicting expected outcomes. Using the same method as above, a regression equation was estimated including a random error term to impute predicted values for missing data.

Table 1

Correlations Among Cognitive Stimulation Scores and Total Home Scores

	Cogsti 88	Cogsti 90	Cogsti 92	Home 88	Home 90	Home 92
Cog Sti 88	1.00					
Cog Sti 90	0.59	1.00				
Cog Sti 92	0.55	0.66	1.00			
Home 88	0.85	0.59	0.54	1.00		
Home 90	0.56	0.86	0.63	0.64	1.00	
Home 92	0.52	0.59	0.86	0.57	0.66	1.00

Level-3 Model: Child Characteristics

In level-3, time invariant child characteristics such as gender, race, the initial test month, and children's verbal memory pretest were used to investigate the effect of each variable on the children's ability growth. In order to investigate the effect of verbal memory on the importance of each sentence characteristic slope, and to investigate the effect of verbal memory on the rate of change of each linguistic/psycholinguistic slope, verbal memory was used for both intercept of each linguistic/psycholinguistic predictor and rate of change slope predictor of each variables. At this level, each parameter (coefficient) from level-2 becomes an outcome:

$$B_{00k} = G_{000} + G_{001}(\text{test month})_k + G_{002}(\text{sex})_k + G_{003}(\text{verbal memory})_k + G_{004}(\text{Hispanic})_k + G_{005}(\text{black})_k + U_{0k}$$

$$B_{01k} = G_{010} + G_{011}(\text{verbal memory})_k$$

$$B_{02k} = G_{020}$$

$$B_{03k} = G_{030}$$

$$B_{10k} = G_{100} + G_{101}(\text{verbal memory})_k$$

$$B_{11k} = G_{110} + G_{111}(\text{verbal memory})_k$$

$$B_{12k} = G_{120}$$

$$B_{20k} = G_{200} + G_{201}(\text{verbal memory})_k$$

$$B_{21k} = G_{210} + G_{211}(\text{verbal memory})_k$$

$$B_{22k} = G_{220}$$

$$B_{30k} = G_{300} + G_{301}(\text{verbal memory})_k$$

$$B_{31k} = G_{310} + G_{311}(\text{verbal memory})_k$$

$$B_{32k} = G_{320}$$

G_{000}: expected ability of a typical child at age 8.5, controlling for gender, race, the initial test month, verbal memory, text characteristics of items, and cognitive stimulation score

G_{001}: effect of the initial test month at age 8.5 on child k's ability, controlling for all the other variables

G_{002}: gender gap in ability at age 8.5, controlling for all the other variables (boys are coded as 1 and girls are coded as 0).

G_{003}: effect of verbal memory at age 8.5 on child *k*'s ability, controlling for all the other variables

G_{004}: adjusted mean ability differences between Hispanic and White children at age 8.5 (Hispanic children are coded as 1 and others are coded as 0), controlling for all the other variables

G_{005}: adjusted mean ability differences between Black and White children at age 8.5 (Black children are coded as 1 and others are coded as 0), controlling for all the other variables

G_{010}: average growth rate in ability at age 8.5, controlling for all the other variables

G_{011}: verbal memory effect on growth rate of child *k*'s ability at age 8.5, controlling for all the other variables

G_{020}: average acceleration of ability, controlling for all the other variables

G_{030}: average effect of home cognitive stimulation score at age 8.5, controlling for all the other variables

G_{100}: average effect of sentence length at age 8.5, controlling for all the other variables

G_{101}: average effect of verbal memory on sentence length at age 8.5, controlling for all the other variables

G_{110}: average linear grow rate effect of sentence length at age 8.5, controlling for all the other variables

G_{111}: average effect of verbal memory on sentence length growth rate at age 8.5, controlling for all the other variables

G_{120}: average acceleration on sentence length, controlling for all the other variables

G_{200}: average effect of vocabulary frequency at age 8.5, controlling for all the other variables

G_{201}: average effect of verbal memory on vocabulary frequency effect at age 8.5, controlling for all the other variables

G_{210}: average linear growth rate effect of vocabulary frequency at age 8.5, controlling for all the other variables

G_{211}: average effect of verbal memory on vocabulary frequency growth rate at age 8.5, controlling for all the other variables

G_{220}: average acceleration on vocabulary frequency, controlling for all the other variables

G_{300}: average effect of propositional density at age 8.5, controlling for all the other variables

G_{301}: average effect of verbal memory on propositional density at age 8.5, controlling for all the other variables

G_{310}: average linear growth rate effect of propositional density at age 8.5, controlling for all the other variables

G_{311}: average effect of verbal memory on propositional density growth rate at age 8.5, controlling for all the other variables

G_{320}: average acceleration on propositional density, controlling for all the other variables

U_{0k}: random effect associated with an individual child *k* at age 8.5, controlling for initial test month, sex, verbal memory, race, and home cognitive stimulation score

Initial test month at level-3 was used to investigate whether there exist any other environmental effect on the assessment. The range of the month of taking the test in 1988 was May to December, with 97 percent of children taking the test between June and October.

Verbal memory, which was assessed around two years before the collection of the PIAT item responses, was used because it has been shown to be a good indicator of children's cognitive development, especially language learning. A study with Spanish children using the McCarthy Verbal memory sub-scale (McCarthy, 1972) showed a moderately high correlation with reading achievement (from r =0.43 to r =0.57). Verbal memory also correlated with the PIAT Reading Recognition (r =0.59) and the PIAT Reading Comprehension (r =0.39). Verbal memory was also correlated (r =0.42) with vocabulary knowledge (PPVT-R), an indicator of verbal intelligence (Baker et al., 1993). In addition, Baddeley et al.'s (1975) study of the effect of articulation on retrieval indicated that the phonological loop in working memory was the key gateway to verbal memory. Older children articulated more rapidly than younger children, and the repetition of words prevented the decay of information from the phonological store. Thus, this articulation speed was directly related to recall. Because many if not most children in this study were in the decoding stage of reading at the beginning of data collection in 1988, this study indirectly investigated the effect of verbal memory on children's reading abilities.

As I indicated in chapter 2, verbal memory was assessed roughly two years before, the PIAT item responses were collected. The correlation between the month of taking the verbal memory and the verbal memory score was low (r=-0.167, n = 448). Imputation for missing value was also conducted by adding randomly generated errors to the mean. The selected verbal memory subtest for assessing the NLSY children is only one part that forms the complete McCarthey assessment battery. Verbal memory was administered by first asking the child to repeat words or sentences said by the interviewer. The child listens to what the interviewer says and retells words or sentences.

There are three parts in the verbal memory subtest: In part A, a child repeats a series of words, ideally in the same sequence. In part B, a child repeats key words. Based on the combined score of parts A and B, Part C--story telling--is administered. Since there are many missing values on part C due to a low score in the combined score of part A and B, I used a standardized combined score of A and B for this study. Verbal memory in the level-3 model was used as an intercept (child's reading ability) predictor. The development of children's reading abilities were observed while controlling for verbal memory along with other intra-individual variables.

Research Questions

1. Do children's reading abilities change at a constant rate?
   a) Do abilities increase at constant or variable rates over time?
   b) Do changes in reading abilities differ across individuals?
2. How does the importance of each linguistic/psycholinguistic variable change as children grow older?

Is the rate of change for each linguistic variable constant or variable?

3. How do individual children's characteristics such as verbal memory interact with text characteristics, such as length of sentence, vocabulary frequency, and propositional density?

   a) Does the effect of sentence length on reading comprehension depend on children's verbal memory?

   b) Does the effect of vocabulary frequency on reading comprehension depend on children's verbal memory?

   c) Does the effect of the propositional density on reading comprehension depend on children's verbal memory?

4. How do contextual factors influence children's growth in reading?

   To what extent does children's growth in reading depend on

   a) verbal memory?

   b) home environment?

   c) race?

   b) gender?

   d) the initial test month?

Summary

By considering items as being nested in occasions, and occasions as being nested in individual child, a three-level HGLM was constructed. The model was used to investigate the patterns of importance of text characteristics along with the patterns of individual child's reading ability. For level-1, such predictors as sentence length, average vocabulary frequency, and propositional density were included. The selection of level-1 predictors was based on information processing theory. To understand the pattern of development over years, three predictors such as age linear, age quadratic, and home cognitive stimulation were also included. In addition, the effect of intra-individual factors on children's reading abilities were investigated. Indirectly, this study investigated the possible source of variance that each cluster of variable explained.

CHAPTER 4

RESULTS

This study was conducted to understand how the characteristics of test items that interact with a child's background shape our beliefs about growth in reading. More specifically, this study was an investigation of the developmental patterns of children's reading abilities and the changing patterns of the importance (effect) of linguistic and psycholinguistic variables in the texts children encountered. In addition, several other factors that might conceivably influence young children's reading abilities, such as characteristics of individuals and characteristics of the contexts in which children learn and develop, were investigated. By building a three-level hierarchical generalized linear model (HGLM), a strong test of this developmental model was possible. The level-1 model represented item characteristics, the level-2 model represented change over time, and the level-3 model represented characteristics of individuals. The following analyses were based on the results for 466 six-year-old children out of 477 who scored more than 15 in the PIAT reading recognition test. Due to missing information on reading comprehension responses, 11 cases were deleted automatically when a three-level HGLM was run.

Patterns of Children's Reading Ability

In order to answer how children's reading abilities change over time, a model (see Figure 2) with both linear and quadratic terms was constructed after building a level-1 model with the three variables, *length* (sentence length), *frequency* (average vocabulary frequency), and *density* (propositional density). Since this study was intended to investigate a typical child's growth pattern over time, the results, which are presented in Table 2, were based on the unit specific model.[4]

Level-1 Model

$\text{Prob}(Y=1|B) = P_{ijk}$

$\log[P/(1-P)] = P_{0jk} + P_{1jk}(\text{length})_{ijk} + P_{2jk}(\text{vocabulary})_{ijk} + P_{3jk}(\text{density})_{ijk} + e_{ijk}$

Level-2 Model

$P_{0jk} = B_{00k} + B_{01k}(\text{age linear})_{jk} + B_{02k}(\text{age quadratic})_{jk} + R_{0jk}$

$P_{1jk} = B_{10k} + B_{11k}(\text{age linear})_{jk} + B_{12k}(\text{age quadratic})_{jk}$

$P_{2jk} = B_{20k} + B_{21k}(\text{age linear})_{jk} + B_{22k}(\text{age quadratic})_{jk}$

$P_{3jk} = B_{30k} + B_{31k}(\text{age linear})_{jk} + B_{32k}(\text{age quadratic})_{jk}$

Level-3 Model

$B_{00k} = G_{000} + U_{0k}$

$B_{01k} = G_{010}$

$B_{02k} = G_{020}$

$B_{03k} = G_{030}$

$B_{10k} = G_{100}$

$B_{11k} = G_{110}$

$B_{12k} = G_{120}$

$B_{20k} = G_{200}$

$B_{21k} = G_{210}$

$B_{22k} = G_{220}$

$B_{30k} = G_{300}$

$B_{31k} = G_{310}$

$B_{32k} = G_{320}$

Figure 2. Patterns of Change in Ability and Change in the Importance of Item Text Characteristics.[5] Notation can be read as item i (n=66), time point j (n=3), and child k (n=466).

[4] The nonlinear HGLM output has two models, the unit specific model and the population average model: The unit specific model incorporates random effect, but the population average model does not. In this study, the interpretation of the results is based on the unit specific model because of the nature of the distribution (non-normal distribution), the average does not reflect a typical child's reading development.

[5] The estimated level-1 variance (1.02069) is close to 1 which indicates little or no over-dispersion. The reliability of the level-1 intercept was 0.416 and the reliability of the level-2 intercept was 0.677. Level-2 and level-3 variances were still significant at p<0.001 level.

Table 2

Non-Linear Model With the Logit Link Function: Unit-Specific Model

Fixed Effect	Coefficient	Standard Error	Approximate T-ratio	df	p-value
Intercept-3, G_{000}	-0.455395	0.026156	-17.411	465	0.000
linear intercept-3, G_{010}	0.025227	0.000612	41.196	734	0.000
quadratic intercept-3, G_{020}	-0.000107	0.000040	-2.719	734	0.007
Length slope, P_1					
intercept-3, G_{100}	-0.079315	0.002759	-28.753	72039	0.000
age linear, G_{110}	-0.002715	0.000086	-31.715	72039	0.000
age quadratic, G_{120}	0.000017	0.000006	3.052	72039	0.003
Vocabulary slope, P_2					
intercept-3, G_{200}	0.095055	0.004152	22.892	72039	0.000
age linear, G_{210}	0.002689	0.000133	20.228	72039	0.000
age quadratic, G_{220}	-0.000052	0.000009	-5.999	72039	0.000
Density slope, P_3					
intercept-3, G_{300}	2.054193	0.199660	10.288	72039	0.000
age linear, G_{310}	0.080118	0.006474	12.376	72039	0.000
age quadratic G_{320}	-0.000572	0.000419	-1.364	72039	0.173

The level-3 intercept G_{000}, which represents children's overall average reading ability across three time points and across items, was statistically significant ($p<0.001$), but the transformed probability was below 0.5. This means that a typical child's average ability of making a correct response was less than 0.5. After controlling for the three linguistic/psycholinguistic predictors, the average reading ability expressed in log-odds was -0.455395,[6] which meant that the probability of making a correct response at a time

[6] To help make sense of the scale reported in this chapter, Appendix shows the value of log-odds and its transformed probability. The value of the log-odds ranges from negative infinity to positive infinity. However, meaningful values of log-odds tend to range from negative 3 to positive 3. Log-odds of 0 represents 0.5 probability (50 percent) of making a correct response. The formula used for transforming log-odds into probability is $p = \dfrac{1}{(1+\exp^{(-(\eta))})}$, where η is the value of the log-odds.

point j by a typical child i was about 0.39. In addition, both linear and quadratic terms were statistically significant. The quadratic effect in the level-1 intercept indicated that a typical child's reading ability did not change at a constant rate (G_{020}= -0.000107 with p<0.001) across the three time points. However, there was a relatively strong linear trend (t = 41.196) compared to the quadratic trend (t = -2.719). The coefficient of the linear trend (G_{010}) was 0.025227 in log-odds, which meant that the reading ability increased over time. The deceleration trend in ability growth over time (G_{020} = -0.000107) suggested that reading ability did not increase as much from 1990 to 1992 as it did from 1988 to 1990. The data in Table 3 and Figure 3 portray this decelerating growth pattern over three time points.

Table 3

Reading Ability by Time

Month (age)	Log-odds
77 (6.4)	-1.125
102 (8.5)	-0.444
125 (10.4)	0.067

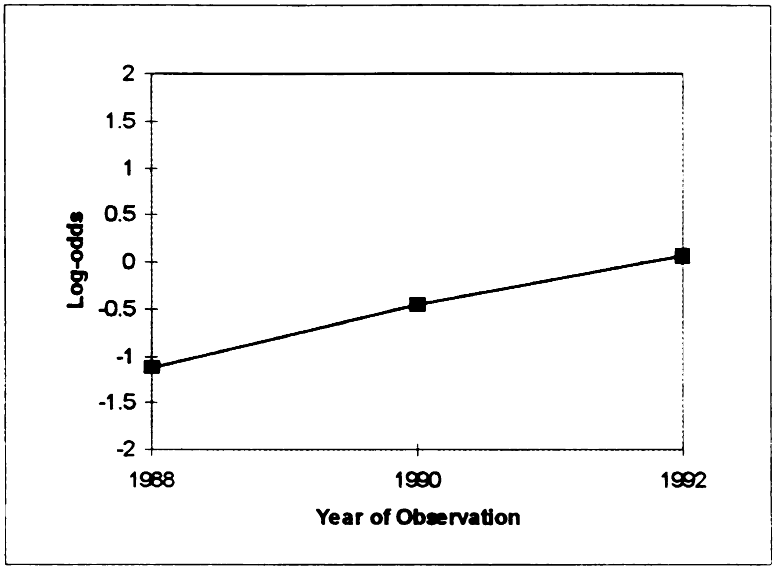

Figure 3. The Growth of Children's Ability in Reading.

Patterns of Linguistic and Psycholinguistic Variables.

In order to investigate whether or not the impact of linguistic/psycholinguistic text variables changes over time, both linear and quadratic terms were used for each of the level-1 slope coefficient predictors. The descriptive statistics for each sentence characteristic are reported in Table 4.

Table 4

Descriptive Statistics for Level-1 Variables

Variable	Mean	Std Dev	Minimum	Maximum	N
Length	14.04	6.58	5.00	31.00	66
Vocabulary	63.67	4.92	49.45	72.70	66
Density	0.34	0.09	0.11	0.67	66

Sentence Length

The average partial effect of sentence length was -0.079315 in log-odds, which meant that the average effect of sentence length on the probability of making a correct response worked negatively. If all the other sentence characteristics are the same, adding one word to the average sentence length makes it more difficult to make a correct response by –0.079315 in log-odds. Then the negative effect of sentence length on the probability of making a correct response increased by -0.002715 (G_{110}) per roughly every two years, but the rate of the increase in performance decreased as children grew older by 0.000017 (G_{120}). This was found in the significant interactions between sentence length and the age terms (G_{110} and G_{120}).

To understand the effect of age with respect to the characteristics of sentence length on the probability of making a correct response in depth, I investigated each of the sentence characteristics further. The impact of time variations on the probability of making a correct response with respect to sentence length is documented in Table 5 and Figure 4. The effect of age on the probability of making a correct response varied depending on the length of a sentence. In the case of a sentence that is one standard deviation shorter than the average, the child's rate of growth was far greater than that with a long sentence--one standard deviation above the average. A child's rate of growth in reading comprehension was almost absent with long sentences (See Figure 4).

Table 5

The Effect of Sentence Length by Time in Log-odds

	Age in Months		
Length	77	102	125
+1sd	-1.1505	-0.974	-0.80
-1sd	-1.0889	0.085	0.96

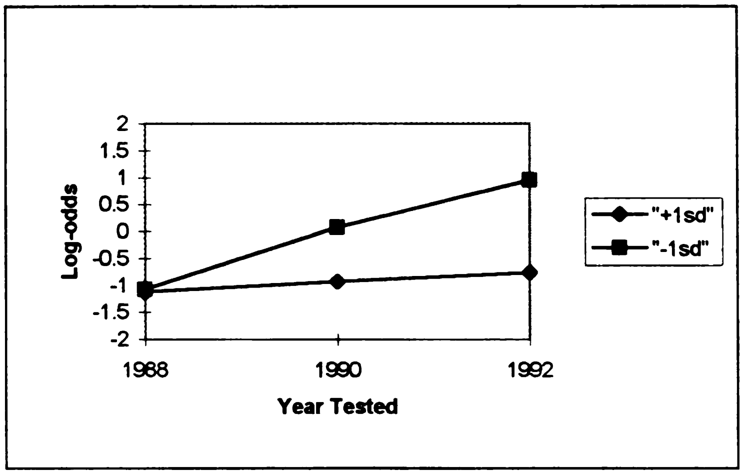

Figure 4. The Effect of Sentence Length.

Vocabulary Frequency

The average effect of vocabulary on the probability of making a correct response was 0.095055. The positive slope (G_{210}= 0.002689) indicated that the effect of vocabulary frequency increased as children grew older. In addition, the negative slope of age quadratic was -0.000052, which was statistically significant at p< 0.001 level. This meant that the effect of the vocabulary frequency increased more between time point 1 and 2 than between time points 2 and 3 (decelerated at each time point).

To understand the effect of time on the probability of making a correct response with respect to characteristics of vocabulary frequency, I investigated items in which the vocabulary frequency was either one standard deviation above or below the average vocabulary frequency. For a sentence in which vocabulary frequency is one standard deviation above the average--that is, a sentence composed of high-frequency words--the growth rate, as reflected in the probability of making a correct response, was larger than that with low frequency words (See Table 6 & Figure 5). Again, this effect was based on the significant interactions between vocabulary frequency and the age terms (G_{210} and G_{220}).

Table 6

The Effect of Vocabulary Frequency by Time in Log-odds

	Age in Months		
Vocabulary	77	102	125
+1sd	-1.1147	0.0295	0.715
-1sd	-1.1248	-0.9176	-0.559

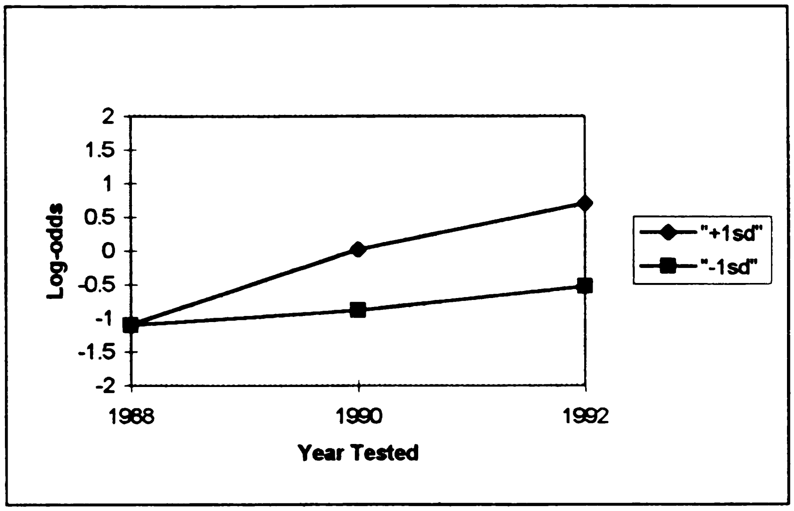

Figure 5. The Effect of Vocabulary Frequency.

Propositional Density

The average effect of propositional density (density = number of propositions in a sentence ÷ number of words in that sentence) was 2.054193 after controlling for other linguistic and psycholinguistic variables. A one unit increase in propositional density would be the difference between a sentence with no propositions and a sentence in which each word was a separate proposition, which in reality could never happen. In this study the values of propositional density ranged from 0.11 to 0.67. It is practically impossible to find a sentence without any proposition and a sentence in which every word is a proposition. Thus, to facilitate understanding of the effect of propositional density, I compared a sentence having low value in propositional density with a sentence having high value. Consider the following:

A windstorm is making a ruin of a cottage.

(# of proposition=1)

(density of propositions = 0.11)

Extremely strong windstorms completely ruined shabby cottages.

(# of propositions =5)

(density of propositions=0 .71).

In this sense, a sentence composed of high content words (meaningful sentence) facilitates reading comprehension.

The rate of change in the linear slope of propositional density was G_{310}= 0.080118, which meant that the positive effect of density on the probability of making a correct response increased over time. The nonlinear growth rate of propositional density was G_{320} = -0.000572, which was not statistically significant. The non-statistical significance of the quadratic effect (G_{320}) of propositional density implied that the rate of increase in the positive effect of propositional density over time was constant.

To understand the effect of time on the probability of making a correct response with respect to characteristics of propositional density, I investigated both one standard deviation above and below the average propositional density. For a sentence in which propositional density is one standard deviation above the average (a highly compact/coherent sentence), the growth rate as was reflected by the probability of making a correct response was greater than that of a sentence one standard deviation below the average. This implied that coherent meaningful sentences (i.e., sentences in which the ideas are packed together) facilitated children's reading comprehension more as the children grew older (See Table 7 & Figure 6). This effect was based on the interaction between density and the linear age term (G_{310}).

Table 7

The Effect of Propositional Density by Time in Log-odds

	Age in Months		
Density	77	102	125
+1sd	-1.1073	-0.2559	0.4319
-1sd	-1.1322	-0.6322	-0.2764

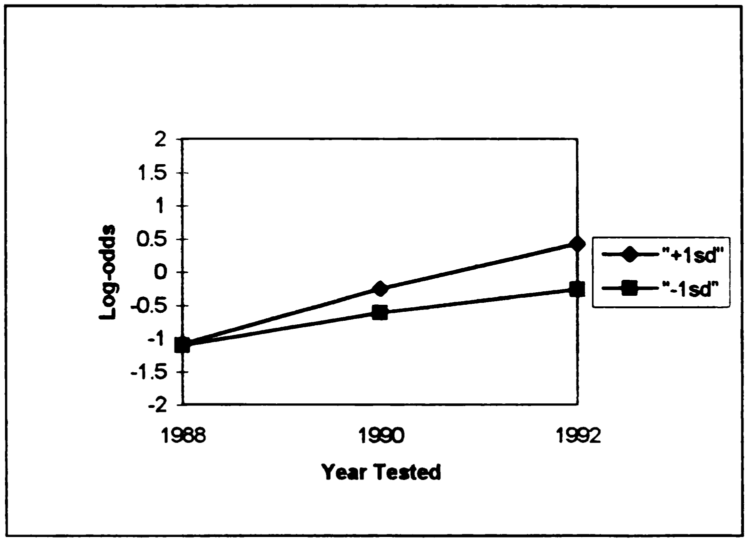

Figure 6. The Effect of Propositional Density.

The effects of the linguistic variables were not constant across time. The effect of sentence length varied by year. At the beginning, in 1988 when the children were around 6.4 years old, the effect of sentence length was minimal. Across time, however, sentences made differential contributions as children grew older. Their comprehension of short sentences increased in a linear fashion while their comprehension of longer sentences did not improve. The same was true with vocabulary frequency. When children were around 6.4 years old, frequency did not make much difference. However, as children grew older, children could understand texts with common words better than those with rare words when the other sentence characteristics were controlled. Ironically, for propositional density, it was with high density sentences that showed increases in comprehension as children grew older. This interpretation must, however, be tempered by the realization that many of the scores of the children at age 6 were near the floor of the test. In other words, when the children were young (at the beginning of data collection in 1988), sentence characteristics did not make much difference in the probability of making correct responses, partly because of the children's limited responses to any of the PIAT items at that time point.

Effect of Contextual Factors on Reading Comprehension

In order to investigate the effect of contextual factors on the growth in reading, variables representing intra-individual characteristics such as gender, race, verbal memory, and the initial test month, were put into the level-3 model as level-2 intercept (ability) predictors. Descriptive statistics for the level-3 variables are in Table 8. In addition, I put home cognitive stimulation score in the level-2 model to look at the effect of changing environmental characteristics on reading ability each time point. The full model, including the full set of factors in level-2, is represented in Figure 7, and the results are presented in Table 9.

Table 8

Descriptive Statistics for Level-3 Variables

Variable	Mean	Std Dev	Minimum	Maximum	N
Sex	0.46	0.50	0	1	477
Test Month	7.97	1.18	5	12	477
Age in Month	77.45	3.59	72	8	477
Verbal memory	95.71	13.73	52	130	477

Note. Sex is a dummy variable, male coded as 1; female coded as 0. The mean for Sex represents the proportion of boys in the sample.

Level-1 Model

$\text{Prob}(Y=1|B) = P_{ijk}$

$\log[P/(1-P)] = P_{0jk} + P_{1jk}(\text{length})_{ijk} + P_{2jk}(\text{vocabulary})_{ijk} + P_{3jk}(\text{density})_{ijk} + e_{ijk}$

Level-2 Model

$P_{0jk} = B_{00k} + B_{01k}(\text{age linear})_{jk} + B_{02k}(\text{age quadratic})_{jk} + B_{03k}(\text{cognitive stimulation})_{jk} + R_{0jk}$

$P_{1jk} = B_{10k} + B_{11k}(\text{age linear})_{jk} + B_{12k}(\text{age quadratic})_{jk}$

$P_{2jk} = B_{20k} + B_{21k}(\text{age linear})_{jk} + B_{22k}(\text{age quadratic})_{jk}$

$P_{3jk} = B_{30k} + B_{31k}(\text{age linear})_{jk} + B_{32k}(\text{age quadratic})_{jk}$

Level-3 Model

$B_{00k} = G_{000} + G_{001}(\text{test month})_k + G_{002}(\text{sex})_k + G_{003}(\text{verbal memory})_k$
$\quad + G_{004}(\text{Hispanics})_k + G_{005}(\text{black})_k + U_{0k}$

$B_{01k} = G_{010} + G_{011}(\text{verbal memory})_k$

$B_{02k} = G_{020}$

$B_{03k} = G_{030}$

$B_{10k} = G_{100} + G_{101}(\text{verbal memory})_k$

$B_{11k} = G_{110} + G_{111}(\text{verbal memory})_k$

$B_{12k} = G_{120}$

$B_{20k} = G_{200} + G_{201}(\text{verbal memory})_k$

$B_{21k} = G_{210} + G_{211}(\text{verbal memory})_k$

$B_{22k} = G_{220}$

$B_{30k} = G_{300} + G_{301}(\text{verbal memory})_k$

$B_{31k} = G_{310} + G_{311}(\text{verbal memory})_k$

$B_{32k} = G_{320}$

Figure 7. Patterns of Change in Ability and in the Importance of Item Text Characteristics.

Table 9

Full Model

Final Estimation of Fixed Effects: (Unit-specific Model)

Fixed Effects	Coefficient	Sandard Error	Approx. T-ratio	d.f.	P-value
Intercept -3, G_{000}	-0.444431	0.024860	-17.878	460	0.000
testmonth, G_{001}	-0.012310	0.016703	-0.737	460	0.461
sex, G_{002}	-0.046791	0.040353	-1.160	460	0.247
verbal memory, G_{003}	0.006075	0.001505	4.036	460	0.000
Hispanics, G_{004}	-0.059223	0.053134	-1.115	460	0.265
Black, G_{005}	-0.108281	0.045211	-2.395	460	0.017
Age linear, G_{010}	0.025006	0.000609	41.066	733	0.000
verbal memory, G_{011}	0.000132	0.000045	2.948	733	0.004
Age quadratic, G_{020}	-0.000120	0.000039	-3.069	734	0.003
Cognitive stimulation, G_{030}	0.000652	0.000106	6.155	734	0.000
Length slope, P_1					
length main effect G_{100}	-0.078679	0.002736	-28.759	72038	0.000
length X verbal memory, G_{101}	-0.000805	0.000123	-6.548	72038	0.000
length X age, G_{110}	-0.002737	0.000085	-32.231	72038	0.000
length X age X verbal, G_{111}	-0.000021	0.000006	-3.358	72038	0.001
length X age^2, G_{120}	0.000015	0.000006	2.711	72039	0.007
Vocabulary slope, P_2					
vocabulary main effect, G_{200}	0.095428	0.004121	23.158	72038	0.000
vocab X verbal memory, G_{201}	0.000612	0.000189	3.231	72038	0.002
vocab X age, G_{210}	0.002716	0.000132	20.589	72038	0.000
vocab X age X verbal, G_{211}	-0.000013	0.000010	-1.375	72038	0.169
vocab X age^2, G_{220}	-0.000052	0.000009	-6.079	72039	0.000
Density slope, P_3					
density main effect, G_{300}	2.064416	0.198118	10.420	72038	0.000
density X verbal memory, G_{301}	0.014566	0.009189	1.585	72038	0.113
density X age, G_{310}	0.081046	0.006425	12.614	72038	0.000
density X age X verbal, G_{311}	-0.000255	0.000475	-0.536	72038	0.591
density X age^2, G_{320}	-0.000573	0.000417	-1.374	72039	0.169

Notes. "Age^2" represents age quadratic effect, and "vocab" represents vocabulary frequency.

Effect of Individual Characteristics on Achievement

The effects of intra-individual characteristics on children's reading achievement were investigated. There was no effect for either gender or the month of the year in which children were tested. However, race was significant. Black children achieved lower scores (log-odds effect size of-0.108281) than did White children. This means that a Black child who was similar to other White children in terms of gender, verbal memory, and test month had -0.552712 in log-odds.

Verbal memory exhibited a statistically significant effect, although its practical significance was not high. While controlling for all the variables in the level-2 and level-3, the average in log-odds was -0.444431 (39.1%). The proportion correct of a child who has one standard deviation above the average in verbal memory was -0.36 in log odds (41%), while that of a child who has one standard deviation below the average in verbal memory was -0.52784 (37 %)[7]. Table 10 shows the effect of verbal memory on reading comprehension.

Table 10

The Effect of Verbal Memory in Log-odds

	Verbal Memory	Log-odds
-1sd	81.98	-0.527
average	95.71	-0.444
+1sd	109.44	-0.361

[7] $=\{-0.444431 \text{ (grandmean)} \pm (0.006075 * 13.73)\}$

Changing Home Environment

The average effect of home cognitive stimulation was statistically significant ($p<0.001$). The coefficient of cognitive stimulation was $G_{030}= 0.000652$, which means when other predictors were controlled (the same value as the grandmean), a one point change in the cognitive stimulation score improved children's average reading ability by 0.000652 in log-odds. The descriptive statistics for level-2 variables are reported in Table 11.

Table 11

Descriptive Statistics for Level-2 Variables

Variable	Mean	Std Dev	Minimum	Maximum	N
Age linear	0.00	19.74	-29.37	34.63	1431
Age quad	389.24	301.54	0.14	1199.24	1431
Cog stim	988.02	155.20	393.00	1432.00	1431

In order to understand the effect further, I examined its impact at both one standard deviation above average and one standard deviation below average of the home cognitive stimulation score (See Table 12). While the average proportion correct was -0.444431 in log-odds (p = 39 %), for a child who is one standard deviation above average in cognitive stimulation score, the log-odds of the proportion correct was -0.34324 (p= 41.5%). For a child who is one standard deviation below average in cognitive stimulation score, the log-odds of the proportion correct was -0.54549 (p= 36.7%)[8].

[8] = {-0.444431 ± (0.000652*155.20)}

Table 12

The Effect of Home Cognitive Stimulation Score

	Cognitive Stimulation	Log-odds
-1sd	832.82	-0.545
Grandmean	988.02	-0.444
+1sd	1143.22	-0.343

Patterns of Growth in Ability

In addition, the patterns of growth in reading ability and the patterns of the importance of the psycholinguistic variables after controlling for such variables as gender, race, verbal memory, the initial test month, and cognitive stimulation scores were investigated. The patterns of ability were similar to the previous model, before the contextual variables were put into the model. The abilities were increased slightly with a slight decrease in standard error, an indication that this model has a better fit. There was a strong linear trend ($t = 41.066$) compared to the quadratic trend ($t = -3.069$). Verbal memory was significant in predicting the linear trend (rate of growth).

Individual differences associated with average ability still existed after controlling for those intra-individual characteristics. The random effect of level-3 intercept variance (0.11512) was statistically significant at $p<0.001$, which means that there were differences among individuals in the average probability of making correct responses even after controlling for gender, race, verbal memory, and the initial test month, and home cognitive stimulation score along with all the previously included sentence characteristics.

Interaction Between Verbal Memory and Text Characteristics Over Time

Sentence length

In order to look at the effect of verbal memory on the growth with respect to sentence length, I examined the log-odds of responses for students who were ± 1 standard deviation on the verbal memory task for sentences that were ± 1 standard deviation on the sentence length metric at each of the three testing points. The effect of verbal memory over time depended on the level of sentence length. The effect of verbal memory appeared only with short sentences. In other words, there was almost no practical effect of verbal memory with long sentences over time. This implied that verbal memory counted more with short-sentences, when all the other text characteristics were controlled. In addition, the effect of high verbal memory on the growth rate slightly declined annually as was reflected by G_{111} (-0.000021). See Table 13 and Figure 8.

Table 13

Interaction Effect Between Verbal Memory and Sentence Length Over Time on Reading Comprehension in Log-odds

Group Characteristics		Year		
Verbal Mem	Length	88	90	92
+	+	-1.131	-0.948	-0.799
-	-	-0.674	0.249	1.359
+	+	-1.152	-0.969	-0.82
-	-	-0.986	-0.063	1.047

Notes. "+" represents one standard deviation above each variable mean and "-" represents one standard deviation below each variable mean.

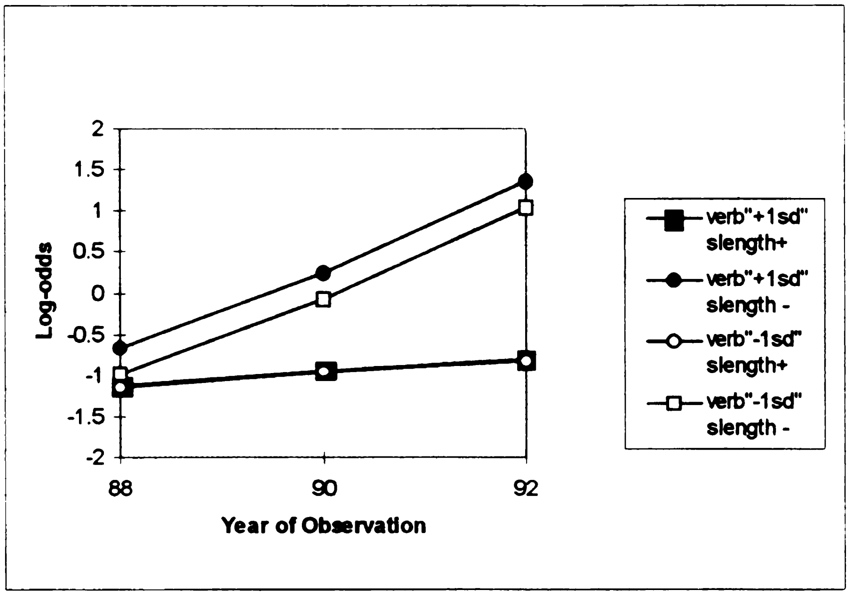

Figure 8. Interaction Effect Between Verbal Memory and Sentence Length over Time on Reading Comprehension in Log-odds.

Vocabulary frequency

In order to look at the effect of verbal memory on growth reflected by vocabulary frequency over time, I examined the log-odds of responses for students who were ± 1 standard deviation on the verbal memory task for sentences that were ± 1standard deviation on the vocabulary metric at each of the three testing points. Verbal memory was statistically significant in predicting the average effect of the vocabulary frequency. In other words, verbal memory effect increased with word frequency. However, the verbal memory on the vocabulary linear growth rate (frequency slope) over time was not statistically significant as was indicated by G_{211}. The effect of vocabulary on the log-odds was greater between 1990 and 1992 than that of 1988 and 1990 as was indicated by G_{210}. See Table 14 and Figure 9.

Table 14

Interaction Effect Between Verbal Memory and Vocabulary Frequency Over Time on Reading Comprehension in Log-odds

Group Characteristics		Year		
Verbal Mem	Vocabulary	88	90	92
+	+	-0.982	0.167	0.843
-	-	-0.935	-0.867	-0.392
+	+	-1.232	-0.082	0.594
-	-	-1.019	-0.951	-0.476

Notes. "+" represents one standard deviation above each variable mean and "-" represents one standard deviation below each variable mean.

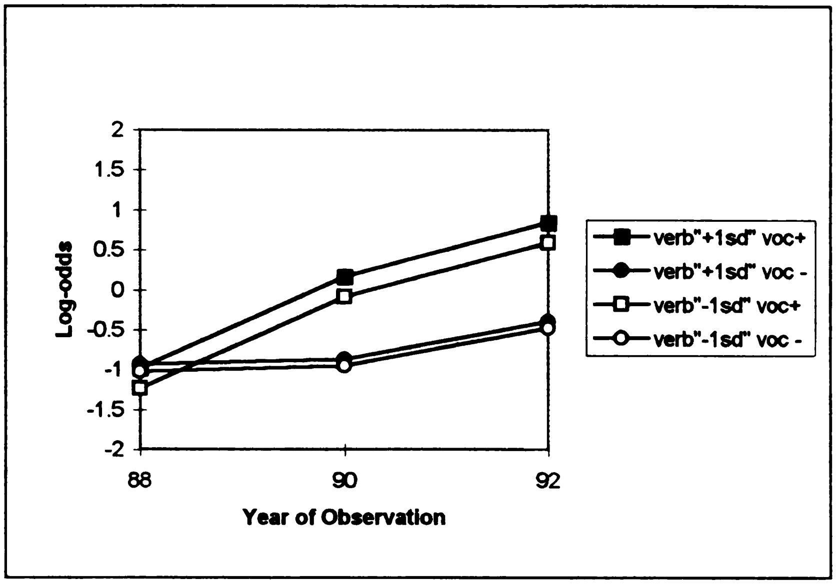

Figure 9. Interaction Effect Between Verbal Memory and Vocabulary Frequency over Time on Reading Comprehension in Log-odds.

Propositional density

In order to look at the effect of verbal memory on the growth reflected by propositional density over time, I examined the log-odds of responses for students who were ± 1 standard deviation on the verbal memory task for sentences that were ± 1 standard deviation on propositional density metric at each of the three points of testing. Verbal memory was neither a significant predictor for the average proposition slope nor a significant predictor for the rate of change over time (linear effect). The consistent rate of change in propositional density was not influenced by the level of verbal memory as was indicated by G_{311}. In addition, the quadratic trend is not statistically significant, which means that the effects of propositional density are consistent over time as was represented by G_{320}. See Table 15 and Figure 10.

Table 15

Interaction Effect Between Verbal Memory and Propositional Density Over Time on Reading Comprehension in Log-odds

Group Characteristics		Year		
Verbal Mem	Density	88	90	92
+	+	-1.028	-0.143	0.507
-	-	-0.89	-0.557	-0.055
+	+	-1.23	-0.346	0.304
-	-	-1.021	-0.688	-0.185

Notes. "+" represents one standard deviation above each variable mean and "-" represents one standard deviation below each variable mean.

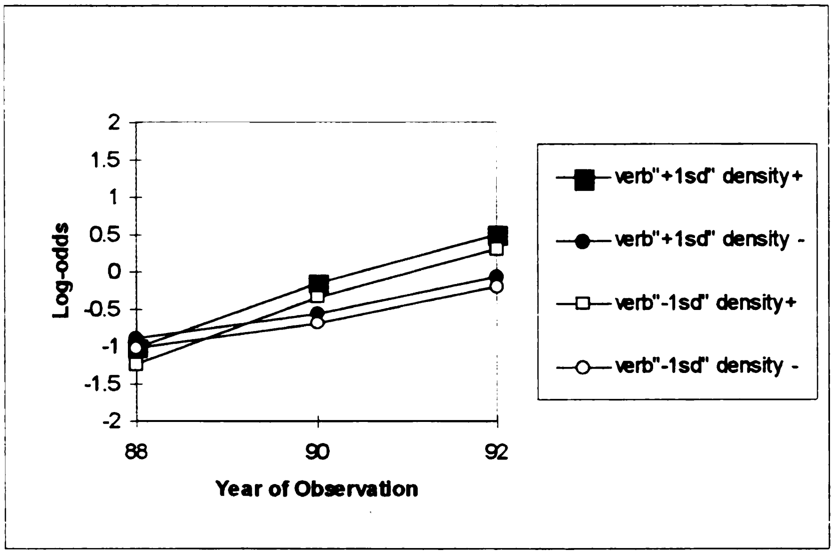

Figure 10. Interaction Effect Between Verbal Memory and Propositional Density over Time on Reading Comprehension in Log-odds.

Summary

Children's reading abilities did not change at a constant rate. However, there was a relatively strong linear trend compared to a quadratic trend. Reading abilities were increasing, but there was a deceleration trend in the growth in abilities over time. To understand the effect of text characteristics on the probability of making a correct response in depth, I investigated each of the sentence characteristics further.

Children's growth in reading comprehension became far larger with short sentences than with long sentences. The effect of time on the probability of making a correct response increased only with short sentences when all the other sentence characteristics were controlled.

The effect of vocabulary frequency increased with time when all the other sentence characteristics were controlled for. There was a deceleration trend over time. The growth rate with frequently used vocabulary was far greater than with less frequently used vocabulary.

The effect of time on the probability of making correct responses increased with high density sentences than with low density sentences when all the other sentence characteristics were controlled for. The rate of increase in the positive importance of propositional density over time was constant. The effect of maturation on the reading comprehension was greater with high density sentences, which are one standard deviation above the mean, than with low density sentences, which are one standard deviation below the mean. This implied that coherent meaningful sentences facilitated children's reading comprehension more as the children grew older

To investigate the effect of contextual factors on children's reading abilities, changing home cognitive stimulation scores were investigated. Home cognitive stimulation scores were statistically significant in predicting reading abilities. Among the four intra-individual characteristics, verbal memory and race were statistically significant.

The interaction effects between verbal memory and text characteristics were also investigated. Verbal memory was statistically significant in predicting both the average effect of sentence length over time and the rate of change of the importance of the sentence length. There was an interaction effect between verbal memory and sentence length. In the case of short sentences, the effect of verbal memory was practically significant. However, in the case of long sentences, the effect of verbal memory was almost absent. Verbal memory was only statistically significant in predicting the average effect of vocabulary frequency. It was not statistically significant in predicting the effect of rate of change of the importance of vocabulary frequency. In the case of propositional density, verbal memory neither predicted the average slope of propositional density, nor did it predict the rate of change at a specific time point.

CHAPTER 5

CONCLUSIONS AND DISCUSSION

Summary

This study investigated in a longitudinal fashion, the patterns of growth in the in beginning readers and the ways in which those growth patterns are influenced by text characteristics. Additionally, individual difference factors, such as gender, race, test month and verbal memory, and one cognitive environmental factor, home cognitive stimulation, were investigated to determine their influence on reading development and their interaction with text variables over time.

The study was simultaneously grounded in three research traditions: information processing theory, readability research, and research on the normal course of early reading development. Of these traditions, readability research is the most relevant to this endeavor. Several researchers (Gray & Leary, 1935; Bormuth, 1964; Draper et al., 1971; and Chall et al., 1990) have claimed that linguistic variables predict comprehension difficulty differentially for subjects at various levels of achievement. In general, this study supports this consistent finding. All three variables--sentence length, vocabulary frequency, and propositional density--are important for children's reading comprehension, but their patterns of influence vary as a function of children's age and reading ability. In addition, this study shows how the influence of each variable changes over time. The effect of time on each text characteristics found in this study do not match the predictions found in the "golden years" of readability research. Rather, this study shows that when ideas are tightly packed together, as they are when propositional density is high, children's reading comprehension is facilitated, and this effect consistently increases as children grow older, when the other text characteristics are the same.

Reading Ability Growth Pattern

In a model that does not account for individual factors, the rate of growth on the PIAT reading comprehension was greater from ages 6 to 8 than from ages 8 to 10. In a successive model, in which individual differences (race, gender, verbal memory, and the initial test month, and home cognitive stimulation score) were used to understand reading achievement, this collection of factors explains a small amount of variation in the growth pattern of the PIAT. In addition, the second model revealed that children with high verbal memory demonstrated greater growth in reading achievement over time than children with low verbal memory. However, even after controlling for the effect of verbal memory on the growth rate, the same overall pattern of growth was shown; that is, the rate of growth from ages 6 to 8 was greater than that of the ages 8 to 10. The rate of growth decreases slightly as children grow older. What the second model demonstrates is that individual differences in the rate of growth exist even after taking account of verbal memory.

Among the intra-individual characteristics, race and verbal memory explain differences in children's reading ability over time, with race demonstrating its usual majority-minority performance differences. The effect of verbal memory on the children's reading ability supports the information processing theoretical roots of this study. The more proficient readers have better memory skills. However, the effect of the month that the children took the test (reading achievement) is not significant, probably due to restricted range. While the potential range is May to December, the effective range is June to October (97 percent of the children took the test in these months).

Psycholinguistic/Linguistic Variable Growth Pattern

As sentence length increases, performance decreases when controlling for all the other item characteristics. For example, items 50 and 70 have equal propositional density index (0.33) and similar average vocabulary frequency index (62.38 vs.62.26), but vary in sentence length (9 vs.21words). As it turns out, there is a substantial difference in performance on these two items, yielding a 20-item differential in placement within the PIAT test.

Item 50.

Occasionally one decides to communicate stealthily with an associate.

Item 70.

In the medieval epoch, a feudal lord often imposed tyrannical punishment on a serf for even the slightest defiance of edicts[9].

The effect of maturation (time) on the children's reading comprehension depends upon sentence characteristics. In the case of sentence length, the effect of time was manifested with short sentence but not with long sentences. From age 6 to 8 to 10, there is virtually no change in students' ability to respond correctly to items with very long sentences; however their capacity to respond correctly to items with short sentences increases across all 4 years, with slightly greater increases from 6 to 8 than from 8 to 10. This differential effect may be an artifact of the difficulty of this test for this population; it may show little more than the fact that the students in this sample, in general, did not do well on the harder (and later) items on this test.

---

[9] In the PIAT reading comprehension test, there do not exist items with the same number of propositions and similar vocabulary frequency value but different sentence length.

In general, items with more frequently used vocabulary are better understood than those with less frequently used vocabulary, when all the other item characteristics are held constant. For example, item number 20 and item number 50 have the same sentence length (9 words) and the same propositional density index (0.33), but the index of vocabulary frequency is different (70.74 vs. 62.38). This difference in vocabulary frequency corresponds to a "30 item" difference in the placement on this test.

Item 20.

It is fun to play with boats that sail.

Item 50.

Occasionally one decides to communicate stealthily with an associate.

The effect of vocabulary frequency on children's reading comprehension increases over time. The pattern of the effect is similar to that for sentence length: Children exhibit little growth in their capacity to respond to sentences with many lower frequency words (one standard deviation below the mean frequency), but they demonstrate a slightly decelarating increase in their capacity to respond correctly to items with higher frequency words (one standard deviation above the mean).

The impact of propositional density does not follow the pattern of performance observed for sentence length and vocabulary frequency. The findings for propositional density seem to contradict the existing readability assumptions (see Pearson, 1974-5). This study shows that the existence of many ideas in a sentence does not work as an hindrance in reading comprehension. Sentences with higher propositional density facilitate children's reading comprehension, if all the other text characteristics are controlled for. For example, items 29 and 37 have the same length (9 words) and similar average vocabulary frequency index (69.26 vs. 69.74), but they have a difference of more than one standard deviation in propositional density index (0.33, vs. 0.22). Due to low propositional density, item 37 can be considered as relatively difficult.

Item 29

The train has a long truck on a flatcar.

Item 37

The purse was on a footstool near the television.

The importance of propositional density increases constantly with age. The growth in reading was far greater with high density sentences than with low density sentences, which means that children understand better when sentences are coherent/compact as they grow older, when all the other sentence characteristics are controlled.

Unlike the other two predictors, length of sentence and vocabulary frequency, the contribution of a large propositional density to reading comprehension in some way contradicts traditional readability studies, which do not consider reading processes in depth. While traditional readability studies get at surface level difficulty, they do not reveal the deeper internal processes in the same way that the propositional density factor does.

The importance of the linguistic variables is not constant over time. In early reading development, the increase in sentence length hinders reading comprehension, while the use of frequently found vocabulary and the use of a compact (coherent) sentence structure facilitate children's reading comprehension. However, it must be noted that for the six-year-old children, none of the sentence characteristics make much difference on

children's reading comprehension, most likely because of children's limited responses to the PIAT items. Since they did not get very far on the test, there was a very restricted range for each of these linguistic variables.

Relationship Between Verbal Memory and Psycholinguistic Variables

Verbal memory tended to emerge as an explanatory factor in interaction with other variables. For example, the average importance of sentence length on reading comprehension depends on verbal memory. In the case of long sentences, verbal memory does not make a difference on children's reading comprehension across the three time points, but its impact on comprehension given short sentences is consistently increasing across the three time points: Students with high verbal memory show a consistent advantage over those with low verbal memory for these shorter sentences. Considering the fact that verbal memory score reflects children's ability to retain information for a certain duration of time, its effect on long-length sentence is limited. The limited contribution of verbal memory to children's reading comprehension may be ascribed to the nature of the verbal memory test. The verbal memory score in this study is based on parts A and B of the McCarthey assessment. Part A measures a child's ability to repeat a series of words in order, part B measures whether or not a child can repeat key words in a sentence, and part C, the part not considered in this study, measures whether they can recall key ideas from that story.[10] In this sense, verbatim retention of words as measured in parts A and B of McCarthey assessment may not relate to the capacity for understanding/ retaining a long sentence.

---

[10] Part C is given to the children whose combined score in the parts A and B is over 8. Since many of children in this study did not obtain a combined scores over 8 in the two parts, verbal memory score for this study was based on only parts A and B.

Vocabulary frequency improves reading comprehension more for children with high verbal memory than for children with low verbal memory. Verbal memory neither predicts the average importance of propositional density, nor does it predict the rate of change in the importance of propositional density.

Discussion

The effects of sentence length and vocabulary are not surprising. Both have an impact on comprehension, as has been demonstrated again and again across the decades. The only twist in this study, that improvements in comprehension are greater for easier (items with high frequency words and shorter sentences) than harder items, may be more a function of the difficulty of this test for this sample of students than anything else. On the other hand, the positive impact of propositional density on children's reading comprehension contradicts the traditional readability studies, which are based on the external characteristics of text. In particular, the fact that high density items are more readily understood than low density items, when all the other sentence characteristics are controlled, reveals more about internal cognitive processes than external features of text. Understanding a text not only depends upon the number and frequency of words, but upon internal coherence of the sentences. Children better understand a text which is composed of coherent, tightly packed meaning units than a text with loosely packed ideas whose interrelationships may have to be inferred by the reader.

The finding that a strong positive importance of propositional density reflected on the large coefficient is more compatible with Pearson's (1969, 1974-5) findings with 3rd and 4th graders' reading processes. Obtaining similar results for conceptually dense sentences, Pearson argues that comprehension consists more of synthesis than analysis.

Hence, more coherently packed sentences are more readily understood because they are closer to the structure in which they will have to be processed in short-and long-term memory, and less inference is necessary to process them compared to low density sentences.

The current study adds strength to this line of work by virtue of several design characteristics; by incorporating a longitudinal dimension and individual characteristics, this study reveals more than the earlier studies did or could. In particular, the longitudinal design permits a careful examination of the rates of changes in each relevant variable. Each of the three variables, sentence length, vocabulary frequency, and propositional density, has a different rate of growth and shows its differential contribution to children's reading achievement over time. Most significant is the fact that the importance of the density variable increases constantly; moreover, the more densely packed sentences are better understood than those that are less dense, except at age 6, when there is no differential effect between high and low density on children's reading comprehension.

Implications for Test Development and Methodology

This study has implications for test development and reading comprehension. Reading comprehension depends on various factors, such as time, text characteristics, and individual characteristics. This study focused especially on the different contribution that each text characteristic makes to the children's reading achievement over time. First, the contribution that each text characteristic makes on the children's reading comprehension changes over time. Second, the contribution that each item text characteristic makes on children's reading comprehension also depends upon the level of each text characteristic. Third, there is a different contribution of verbal memory on the rate of change in the importance of text characteristics (sentence length) over time, which implies that the manifestation of intra-individual characteristics also depends upon text characteristics. Thus, high achievement in a particular reading comprehension test does not imply that the children's abilities improved as such. This study implies that children's measured reading achievement is influenced by the text characteristics of the items in an achievement assessment instrument.

For the PIAT item writers and for any test developers who want to build a multi-age appropriate test in which the items are ordered according to difficulty, this study implies that the rank order of item difficulties can be obtained by manipulating these three linguistic/psycholinguistic variables in the test design stage. Clearly, test developers know the effects of sentence length and vocabulary frequency. The real news for them is likely to be the impact of the propositional density variable.

In addition, this study implies that the effect of each text characteristics is also different with different ability groups. In developing age-appropriate texts, it is appropriate for the text writers to consider the changing importance of text characteristics at each time point. In other words, for the authors of trade books, basal readers, and other children's material, this study indicates that if children are exposed to coherent, meaningful texts in addition to short-length sentences and sentences with high vocabulary frequency, children's comprehension may improve as they grow older. When publishers construct texts, considering the change of importance of each text characteristic may result in texts that induce effective learning or information retention (afferent reading). In addition, children's appreciation of literature might also be facilitated.

It is not clear how these results might impact the work of classroom teachers, except perhaps to advise them to examine conceptual coherence as well as "traditional" indicators of readability, like sentence length and vocabulary frequency, when selecting books for their students to read. Based on the current results, even six-year-olds seem likely to be able to handle densely packed texts.

Limitations

There are some limitations in this study. Due to the restriction of age, the follow-up study of six-year-old children in 1988 up to roughly 10 in 1992, the findings cannot be generalized beyond these age groups. The results of this study may be generalized for test items of one sentence length, which excludes most of the tests in the current elementary testing market place. In addition, due to many missing values, this analysis required the imputation of missing values. Even though I used rigorous methods in imputing the missing values, the results may not be exactly the same if I had analyzed the data based on actual responses without missing values. Since the PIAT text items are highly controlled and somewhat artificially constructed, some apprehension related to validity exists. In addition, the number of propositions identified in each sentence might be different if I had used a different method of proposition identification.

Direction for Future Research

Although this study incorporates theoretically meaningful variables at each level, unexplained individual differences at certain time points and across all three time points exist. In addition, in spite of the choice of meaningful variables at level-2 (time points) and level-3 (child characteristics), a significant random effect persists, which implies that the included variables do not explain all of the variance in children's reading achievement.

Since the major focus of this research is looking at the patterns of importance of psycholinguistic variables along with the patterns of children's reading abilities, I did not put much emphasis on incorporating SES-related variables found in the NLSY data. In addition, the data suggest some probable multicollinearity problems within and between levels, if I were to incorporate some SES-related variables. Since I did not incorporate many SES-related variables, much of the variance at level-3 remained unexplained. In the future, the use of more meaningful variables related to intra-individual characteristics is suggested to explain the development of reading abilities.

In addition, this study does not exclude the possibility that other meaningful variables could be incorporated into the model. In the case of level-1 (text characteristics), although there is no evidence of over-dispersion, it may be possible to incorporate other meaningful variables and investigate their importance over time.

This study looked at the growth pattern of six-year-old children's reading comprehension while incorporating text characteristics at two-year intervals from 1988 to 1992. Thus, the reading comprehension of children in 1989 and 1991 were not exactly estimated. If a researcher were to take annual, rather than biannual measures of the PIAT, the effects implied in the current study could be evaluated more precisely. On the other hand, it is possible, using the current NLSY data base to estimate these "between year" performance points. In the 1988 NLSY data base, measures of other ages groups such as 7, 8, 9 were obtained. Hence, the sample of students who were age 7 in 1988 could be used to estimate the performance of students ages 7, 9 and 11. Such analyses may well corroborate the current findings that the effect of the text characteristics depends upon the age/ability of a child, but do so more reliably, by filling in the age gaps missing in the current work and therefore increasing the generalizability of this study.

Improved reading assessment instruments might enhance the reliability of the current results. If the PIAT administration were more consistent with guidelines and if interviewers' coding mistakes were reduced, the reliability of this study would increase. Therefore, the inconsistent administration procedures and the coding mistakes also influence the validity of reading comprehension assessment. It would also be interesting and important to examine other measures of comprehension, preferably measures which are not quite so formulaic as the PIAT, to determine whether the current findings extend to more normal texts, the kind of books children read on an everyday basis.

APPENDIX

APPENDIX

The value of the log odds ranges from negative infinity to positive infinity. However, meaningful values of log odds tend to range from negative 3 to positive 3. Log-odds of 0 represents 0.5 probability (50 percent) of making a correct response.

Possible Ranges of Log-odds vs. Probability

Log-odds	Probability
3	0.95
2	0.88
1	0.73
0	0.50
-1	0.27
-2	0.12
-3	0.05

Note. The transformation between log-odds and probability of making a correct response was not linear.

BIBLIOGRAPHY

BIBLIOGRAPHY

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Anderson, R. C.,& Pearson, P.D. (1984). A schema-theoretic view of basic processes in reading. In P. D. Pearson (Ed.), *Handbook of Reading Research*. New York: Longman.

Baddeley, A. D. , Grant, S. E., & Thomson, N. (1975). Imagery and visual working memory. In P.M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance,* (pp. 205-217). New York: Academic Press.

Baddeley, A. D.& Hitch, G. J. (1974). Working memory. In G. Bower (Eds.), *The psychology of learning and motivation,* (Vol. 8, pp. 47-90). New York: Academic Press.

Baker, P.C., Keck, C. K., Mott, F. L., & Quinlan, S. V. (1993). *NLSY child handbook. A guide to the 1986-1990 national longitudinal survey of youth child data.* (Revised Ed.). Columbus, OH: Center for Human Resource Research, The Ohio State University.

Bormuth, J. R. (1964). *Relationships between selected language variables and comprehension ability and difficulty*. Cooperative Research Project, Number 2082. Los Angeles: University of California.

Bormuth, J. R. (1966). Readability: A new approach.. *Reading Research Quarterly, 1,* 79-132.

Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychological Review, 64,* 205-215.

Bryk, A., & Raudenbush, S. (1992*). Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Bryk, A.S., Raudenbush, S. W., & Condon, R. T. (1996). *Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L program.* Chicago: Scientific Software International Inc.

Carroll, J. B. (1972). Defining language comprehension: Some speculations. In J. B. Carroll and R. O.Freedle (Eds.), *Language comprehension and the acquisition of knowledge*. New York: V. H. Winston & Sons.

Carver, R. P. (1977-1978). Toward a theory of reading and comprehension and rauding. *Reading Research Quarterly, 1*, 9-89.

Carver, R. P. (1981). *Reading comprehension and rauding theory.* Springfield, IL: Charles C. Thomas.

Center for Human Resource Research. (1997). *1994 NLSY '79 child and young adult data user guide.* Columbus, OH: The Ohio State University.

Chall, J. S. (1970). Interpretations of the Results of Standardized Reading Test. In R. Farr (Ed.), *Measurement and evaluation of reading.* (pp. 51-59). New York: Harcourt, Brace & World, Inc.

Chall, J. S. (1983). *Stages of reading development.* New York: McGraw-Hill Book Company.

Chall, J. S., Jacobs, V. A., & Baldwin, L. E. (1990). *The Reading crisis: Why poor children fall behind.* Boston: Harvard University Press.

Crawford, W. J., King, C..E., & Brophy, J. E. (1975). Error rates and question difficulty related to elementary children's learning. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

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

Davis, F. B. (1944). Fundamental factors of comprehension in reading. *Psychometrica, 9*, 185-197.

Davis, F. B. (1972). Psychometric research on comprehension in reading. *Reading Research Quarterly, 7*, 628-678.

Draper, A. G. & Moller, G. H. (1971). We think with words (therefore, to improve thinking, teach vocabulary). *Phi Delta Kapppen, 52,* 482-484.

Dunn, Lloyd M. & Frederick C. Markwardt Jr. (1970). *Peabody individual achievement test: Subtest, reading comprehension.* Circle Pines, MN: American Guidance Service, Inc.

Flesch, R. F. (1943). *Marks of readable style: A study in adult education.* New York: Bureau of Publications, Teachers College, Columbia University.

Gardner, H. (1987). *The mind's new science. A history of the cognitive revolution.* U.S. A.: Basic Books.

Gates, A. I. (1935). *The improvement of reading: A program of diagnostic and remedial methods.* (Rev. ed.) New York: the Macmillan Company.

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

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. (1991). The influences of number of syllables and word-likeness on children's repetition of nonwords. *Applied Psycholinguistics, 12,* 349-367.

Gray, W. S., & Leary, B. A. (1935). *What makes a book readable.* Chicago: University of Chicago Press.

James, W. (1890). *The principles of psychology.* New York: Henry Holt.

Jorm, A. F. (1983). Specific reading retardation and working memory: A review. *British Journal of Psychology, 74,* 311-342.

Harris, A. J. & Sipay, E. R. (1990). *How to increase reading ability: A guide to developmental and remedial methods.* (9th edition). New York: Longman.

Kintsch, W. (1974). *The Representation of meaning in memory.* New York: John Wiley & Sons.

Kintsch, W. (1979). On modeling comprehension. *Educational Psychology, 14,* 3-14.

Kintsch, W., & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L-G.Nilsson (Ed.), *Perspectives on memory research.* Hillsdale, N. J.: Erlbaum.

Klare, G. R. (1963). *The Measurement of readability.* Ames, Iowa: The Iowa State University Press.

Klare, G. R. (1984). Readability. In P. D. Pearson, Barr, M. L. Kamil, & P. Mosenthal (Eds.) *Handbook of reading research.* New York: Longman.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293-323.

Luster, T. & Dubow, E. (1992). Home environment and maternal intelligence as predictors of verbal intelligence: A comparison of preschool and school age children. *Merrill-Palmer Quarterly, 38* (2), 151-175.

Mehler, J. (1963). Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior, 2,* 346-351.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review 63,* 81-97.

Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist, 17*, 748-62.

Miller, P. H. (1993). *Theories of developmental psychology*. (3rd ed.). New York: W. H. Freeman and Company.

Morrison, F. J., Giordani, B., & Nagy, J. (1977). Reading disability: An information processing analysis. *Science, 196*, 77-79.

National Center for Education Statistics (Ed.). (1996). *Results from the NAEP 1994 Reading assessment-at a glance.* Washington, D. C.

Nunnally, J. (1978). *Psychometric theory.* New York: Mcgraw Hill Book Company.

Office of Policy and Planning. U. S. Department of Education. (1992). *Transforming American education; A directory of research and practice to help the nation achieve the six national education goals.*

Ogle, L. T., N. Alsalam, & G. T. Rogers. (1991). *The condition of education,*. Washington, D. C.: U. S. Department of Education, National Center for Education Statistics.

Pearson, D. P. (1969). *The effect of grammatical complexity on children's comprehension, recall and connection of semantic relations.* Unpublished dissertation. Minneapolis, MN.: University of Minnesota.

Pearson, D. P. (1974-75). The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. *Reading Research Quarterly, 2*, 153-192.

Plomin, R., & Daniels, D. (1987). Why are children in the same family so different from each other? *Behavioral and Brain Science, 10*, 1-16.

Posner, M. I. (1982). Cumulative development of attentional theory. *American Psychologist, 37*, 168-179.

Scarborough, H.S. (1998). Early identification of children at risk for reading disabilities; Phonological awareness and some other promising predictors. In *Specific reading disability: A view of the spectrum*, B. K Shapiro, P. J. Accardo, and A. J. Capute, eds. Timonium, MD: York Press.

Squires, D. A., Huitt, W. G., & Segars. J. K. (1983). *Effective schools and classrooms. Alexandria*, VA: Association for Supervisor and Curricular Development.

Stenner, A. J. (1997). The objective measurement of reading comprehension. In response to technical questions raised by California Department of Education Technical Study Group.

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12,* 242-248.

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. I. Posner (Ed.), *Foundations of cognitive science*. (pp. 527-580). Cambridge, Massachusetts: MIT Press.

Yngve, V. H. (1962). Computer programs for translation. *Scientific American 206.* 68-76.

Zakaluk, B. L., Samuels, S. J., and Taylor, B. *(1986).* A simple technique for estimating prior knowledge: Word association. *Journal of Reading, 30*, 56-60.

Zeno, S.M., Ivens, S. H., Millard, R. T., & Rajduvvuri. (1995). *The Educator's word frequency guide*. U.S. A.: Touchstone Applied Science Associate Inc.