



This is to certify that the  
dissertation entitled

DIFFICULT TEXTS AND THE STUDENTS WHO CHOOSE THEM:  
THE ROLE OF TEXT DIFFICULTY IN SECOND GRADERS' TEXT  
CHOICES AND INDEPENDENT READING EXPERIENCES

presented by

Juliet L. Halladay

has been accepted towards fulfillment  
of the requirements for the

Ph.D.

degree in

Curriculum, Teaching, and  
Educational Policy

*Nell K. Duke*

Major Professor's Signature

*September 1, 2008*

Date

MSU is an Affirmative Action/Equal Opportunity Employer

**DIFFICULT TEXTS AND THE STUDENTS WHO CHOOSE THEM:  
THE ROLE OF TEXT DIFFICULTY IN SECOND GRADERS' TEXT CHOICES AND  
INDEPENDENT READING EXPERIENCES**

**By**

**Juliet L. Halladay**

**A DISSERTATION**

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

**DOCTOR OF PHILOSOPHY**

**Curriculum, Teaching, and Educational Policy**

**2008**

## **ABSTRACT**

### **DIFFICULT TEXTS AND THE STUDENTS WHO CHOOSE THEM: THE ROLE OF TEXT DIFFICULTY IN SECOND GRADERS' TEXT CHOICES AND INDEPENDENT READING EXPERIENCES**

By

Juliet L. Halladay

This dissertation describes a study of the relationships between text difficulty, reading comprehension, and reading motivation for a sample of second grade students (n = 70). The study was designed to explore and describe the reading experiences of students who chose to read texts that would commonly be considered too difficult to be read independently. In particular, the study sought to find out more about students' reasons for choosing difficult texts, their comprehension of those texts, and their affective experiences with reading the difficult texts they chose for themselves.

The study focused on students' reading during self-selected, independent reading time in the general education classroom. Five second grade classes participated in the study. Data sources included group assessments of reading ability and reading motivation, student logs of reading choices, individual assessments of reading comprehension and oral reading accuracy, and student interviews. Parallel measures based on students' experiences reading independent-level texts were used as points of comparison. A combination of quantitative and qualitative analyses were used to test existing hypotheses; explore relationships between reading comprehension, text difficulty, and motivation to read; and construct some illustrative descriptions of students' reading experiences with their chosen texts.



One important finding was that the standard criterion of oral reading accuracy, originated by Betts (1946) and advocated by others, is too high for second grade readers. Another finding was that oral reading accuracy and reading comprehension were not consistently related, showing a statistically significant but weak correlation. The results of this study also suggest that the term “frustration level” may be somewhat of a misnomer, as no relationship was found between a student’s oral reading performance and her enjoyment of a text. Half of the students in the sample were identified as having chosen at least one frustration-level text during the data collection period, indicating that the practice of choosing difficult texts is quite common. The students who chose frustration-level texts were more likely to be struggling readers, but they did not differ significantly from students who did not choose frustration-level texts in terms of perceptions of themselves as readers or beliefs about the value of reading. Another finding was that students rarely mentioned the perceived difficulty of a text as a reason for choosing it. Students’ perceptions of text difficulty were relatively accurate, although students were much more aware of their difficulties reading individual words than they were of their difficulties understanding the text as a whole. Finally, students’ enjoyment of texts was not statistically significantly related to their perceptions of text difficulty.

The results of this study have some important implications. They point toward the need for a reevaluation of the assessment methods and placement criteria that are commonly used to place students in texts by matching student reading ability with text readability. This study also suggests a need to think more critically about the concept of text difficulty as a construct used both in research and in practice.

To my parents, Joseph and Carolyn Dickmann –  
for everything, which has been considerable

## ACKNOWLEDGMENTS

There are many individuals who have contributed in some way to this dissertation, both to the study and to the final written product. I must begin by acknowledging all of the teachers and students who participated in the project. The five teachers gave generously of their time, and I enjoyed getting to know each of them. Their dedication to their students and their commitment to quality teaching are admirable. The students are at the heart of this work; it is their ideas and behaviors and beliefs that are under the spotlight, and I am grateful for their willingness to read and talk with me. I thoroughly enjoyed my time with them, and their enthusiasm for reading and learning served as a valuable reminder of what our work as teachers and researchers is all about.

The work of my dissertation committee members also merits a formal thank you. Suzanne, my advisor throughout my years as a doctoral student, has been a steadfast friend and a reliable source of kind words, honest feedback, and chocolate. I am grateful for her incredible generosity, kindness, and wisdom. Nell, my dissertation director, has contributed greatly to my growth as a researcher, writer, and thinker, and I appreciate her willingness to involve me as a colleague on research studies, conference presentations, and writing projects. With her seemingly endless stores of energy, good humor, and intellect, Nell is a challenging and inspiring mentor and a joy to work with. Susan helped me talk through potential research questions and made valuable comments on my dissertation proposal. Throughout my time as a graduate student, I have benefited significantly from her thoughtful feedback, steady encouragement, and kind understanding. Although he was a later addition to my committee, Jere made important

contributions to the design of this study. His ability to work equally well with big ideas and small details was a great help to me in designing a study that was both theoretically and methodologically sound. He was very generous with his time and ideas, and I always left our discussions with my mental gears churning.

Faculty members outside of my committee also contributed to this work. Mark Conley and Mary Lundeberg were helpful in the initial planning stages, helping me focus my research questions and brainstorm methodological possibilities. Janine Certo was an unofficial committee member in many ways, always willing to talk through research ideas and give feedback on my written work. Her calm demeanor and consistent good cheer buoyed my spirits on a number of occasions. And I would certainly be remiss if I did not acknowledge Michael Pressley, who had a tremendous influence on me during the two years that I was fortunate enough to work with him. I often wish I could talk with him one more time, to have him lean back in his chair, with his feet on the table and his hands behind his head, and remind me to “focus like a laser” on the work at hand. As the director of my practicum study, which was in many ways a precursor to this dissertation, Mike was always willing to share his unparalleled expertise and to give assistance when necessary, but I also learned a great deal through his insistence that I do the work myself. He was truly an amazing individual and is sorely missed.

My graduate student colleagues have also been important to me and to my work over the years. Alison Billman and I worked through the process together every step of the way, keeping each other going and defending our dissertations on the same afternoon; I could not be more proud of Alison and everything she has accomplished. Karen Ames, Blakely Tsurusaki, and Lauren Fingeret served as an essential support group throughout

my time at MSU, and their friendship is as valuable to me as any of the other things I have gained during my time in graduate school. A number of other office mates and literacy partners in crime have each been important in their own ways: Katie Hilden, Annie Moses, Kate Roberts, Meagan Shedd, Erin Wibbens, Becky Norman, Nicole Martin, Shenglan Zhang, and Han Park. I value their friendship, and I look forward to working with them as colleagues in the future. In the actual implementation of the study, Laura Jimenez provided some important assistance with interrater reliability. And I simply could not have completed this work without the indispensable assistance of Autumn Dodge. She generously gave a considerable amount of her time to the project – visiting schools, working with students, scoring assessments, and helping with interrater reliability – and I am very grateful for her help.

And a final thank you goes out to my family. My parents, Joe and Carolyn, continue to stand by me in everything I do, and their immense pride in all of their children's accomplishments means the world to me. My in-laws, Kim and Jeanie, have also been incredibly supportive of us during our MSU years. Their willingness to assume additional grandparent responsibilities was invaluable during the final weeks of writing, and our children loved their extra time with them. Our children – Daniel, Anna, and Vivian – helped to make graduate school simultaneously more challenging and more enjoyable, reminding me that my first responsibility is always to my family. And last but far from least, to my husband Patrick, whom I appreciate more than words can possibly say. But if I had to try, I would thank him for his honesty, energy, and intellect, and for his unending faith in me. We are partners in everything we do, and neither of us could have succeeded without the support of the other.

## TABLE OF CONTENTS

LIST OF TABLES .....	xi
LIST OF FIGURES .....	xii
CHAPTER 1	
INTRODUCTION .....	1
CHAPTER 2	
LITERATURE REVIEW .....	7
Overview .....	7
▪ Betts' Reading Level Framework .....	9
▪ Influence of Betts' Framework .....	11
▪ Critiques and Revisions .....	12
Research Base .....	14
▪ Question 1: Do the students who choose texts at their frustration level fit into certain profiles based on gender, motivation to read, or reading ability? .....	14
▪ Question 2: What reasons or purposes do students give for choosing frustration-level texts for independent reading? .....	16
▪ Question 3: What, if anything, do students understand from reading self-selected, frustration-level texts (as compared to independent-level texts) independently? .....	18
▪ Question 4: What are students' perceptions of the difficulty of their chosen, frustration-level texts? .....	24
▪ Question 5: What are students' affective experiences with reading these difficult texts independently? In other words, are self-selected, frustration-level texts actually "frustrating," or are students able to enjoy them? .....	25
▪ Summary .....	28
Theoretical Framework .....	29
▪ Reading Comprehension .....	29
▪ Text Difficulty .....	32
▪ Motivation to Read .....	36
▪ Summary .....	40
CHAPTER 3	
METHODS .....	41
Overview .....	41
Study Context and Sample Description .....	42
▪ Grade Level .....	42
▪ Classroom Context .....	43
▪ Sample .....	44

▪ Classroom Libraries .....	47
Data Collection .....	51
▪ Reading Ability .....	51
▪ Reading Motivation .....	53
▪ Reading Logs .....	54
▪ Identifying Occasions .....	55
▪ Oral Retellings .....	56
▪ Running Records .....	57
▪ Comprehension Questions .....	60
▪ Interviews .....	62
▪ Identifying Matched Occasions .....	63
Data Analysis .....	64
▪ Scoring .....	64
▪ Interrater Reliability .....	70
▪ Identifying the Subsample of Frustration-level Choosers .....	71
▪ Analysis .....	76
 CHAPTER 4	
RESULTS .....	82
General Results .....	82
▪ Reading Ability .....	82
▪ Motivation to Read .....	83
▪ Interactions: Reading Ability, Motivation to Read, and Gender .....	85
▪ Texts .....	88
Question 1: Do the students who choose texts at their frustration level fit into certain profiles based on gender, motivation to read, or reading ability? .....	91
▪ Reading Ability .....	91
▪ Motivation to Read .....	93
▪ Gender .....	94
▪ Question 1: Summary .....	95
Question 2: What reasons or purposes do students give for choosing frustration-level texts for independent reading? .....	96
▪ Overview of Reasons for Text Choices .....	97
▪ Reasons for Choosing Frustration-level Texts .....	99
▪ Question 2: Summary .....	107
Question 3: What, if anything, do students understand from reading self-selected, frustration-level texts (as compared to independent-level texts) independently? .....	108
▪ Oral Reading Accuracy and Comprehension .....	108
▪ Comprehension of Frustration- and Independent-level Texts ...	110
▪ Question 3: Summary .....	111
Question 4: What are students' perceptions of the difficulty of their chosen, frustration-level texts? .....	112
▪ Overall Perceptions of Text Difficulty .....	113

▪ Perceptions Versus Performance .....	114
▪ Question 4: Summary .....	120
Question 5: What are students' affective experiences with reading these difficult texts independently? In other words, are self-selected, frustration-level texts actually "frustrating," or are students able to enjoy them? .....	121
▪ Overall Ratings of Enjoyment .....	121
▪ Enjoyment and Performance .....	122
▪ Question 5: Summary .....	126
<b>CHAPTER 5</b>	
<b>DISCUSSION</b> .....	128
Betts' Criteria: Too Stringent for Second Grade Readers? .....	128
Oral Reading and Comprehension: Uncertain Relationships .....	131
The Frustration Level: A Misnomer? .....	134
Frustration-Level Text Choices: Frequent and Widespread .....	136
Text Difficulty: A Secondary Consideration .....	140
Text Difficulty and Reading Performance: Mixed Results .....	143
Limitations .....	144
Implications .....	147
Future Research .....	150
<b>APPENDICES</b> .....	155
Appendix A: Student Reading Log .....	156
Appendix B: Comprehension Question Framework and Guidelines .....	157
Appendix C: Sample Comprehension Questions .....	158
Appendix D: Student Interview Protocol .....	159
Appendix E: List of Provided Texts .....	160
Appendix F: Oral Retelling Rubric .....	161
Appendix G: Running Record Score Sheet .....	162
Appendix H: Running Record Error Scoring Matrix and Notes .....	163
Appendix I: Running Record Scoring Manual .....	166
Appendix J: Leveling Criteria Matrix .....	168
Appendix K: Coding Manual for Reasons for Text Choices .....	169
Appendix L: Children's Books Cited .....	172
<b>REFERENCES</b> .....	174



## LIST OF TABLES

Table 3.1. Race, socioeconomic status, and achievement data by school .....	46
Table 3.2. Sample distribution across participating classrooms .....	47
Table 3.3. Oral reading accuracy differences and mean scores for frustration-level and independent-level texts, chosen by the subsample of students who chose to read frustration-level texts .....	75
Table 4.1. Reading ability for the whole sample, from the GMRT-4 .....	83
Table 4.2. Motivation to read scores for the whole sample, from the MRP.....	84
Table 4.3. Pearson correlations between GMRT-4 and MRP subtests and totals ..	86
Table 4.4. T-tests of reading ability scores, by gender .....	87
Table 4.5. T-tests of motivation to read scores, by gender .....	88
Table 4.6. Assessed texts by Lexile level and approximate grade level equivalent	90
Table 4.7. T-tests of reading ability scores, by subsample membership .....	92
Table 4.8. T-tests of motivation to read scores, by subsample membership .....	94
Table 4.9. Most common reasons for choosing frustration- and independent-level texts.....	103
Table 4.10. Text difficulty ratings, from student reading logs .....	113
Table 4.11. Ratings of text difficulty by frustration- and independent-level texts .	115
Table 4.12. Difficulty ratings of frustration-level texts, by number of prior readings .....	119
Table 4.13. Ratings of reading enjoyment, from student reading log entries .....	122
Table 4.14. Text enjoyment ratings for the subsample of frustration-level choosers, by text difficulty group .....	123
Table 4.15. Text enjoyment ratings and comprehension scores for the 70 texts chosen by the subsample of frustration-level choosers .....	124

## LIST OF FIGURES

Figure 4.1. Histogram showing distribution of MRP combined scores .....	85
Figure 4.2. Gender breakdown of the total sample and of the two subsamples: frustration-level choosers and non-frustration choosers .....	95
Figure 4.3. Responses to the interview question, “Were there words that you didn’t know?” by text difficulty .....	117
Figure 4.4. Responses to the interview question, “Were there parts of the book/magazine/other that you didn’t understand?” by text difficulty.....	118

## CHAPTER 1: INTRODUCTION

In the primary grades, the issue of matching texts to readers is of particular concern for teachers. As early readers move from understanding letter-sound relationships to fluent reading, teachers seek to provide their students with texts that will support their reading development. For teachers, one of the primary considerations in making these student-text matches is whether or not the text is written at a readability level that is aligned with the student's level of reading skill. In fact, Chall and Conard (1991) found that a strong majority of the elementary teachers they surveyed regarded suitable reading level as the most important consideration in selecting texts for their students. Putting this belief into practice, many primary grade teachers base their instruction around sets of leveled texts, which have been arranged along a continuum of increasing difficulty, as determined by readability formulas and leveling rubrics (Hoffman, Roser, Salas, Patterson, & Pennington, 2000; Mesmer, 2006). Then when students are allowed to choose their own texts to read independently, teachers often guide them toward texts that are at or slightly above their level of reading achievement (Chall & Conard, 1991). The result is that whether the texts are being used for teacher-led instructional purposes or for student-selected, independent reading, they are often matched carefully to individual readers based on determinations of text readability and student reading ability. It is important to note here that this is an oversimplification and not an entirely fair depiction of the varied purposes – including strategy instruction, recreation, socialization, skill practice, and knowledge building – for which many teachers daily use a broad range of texts with their students in order to produce “not just decoders but literate beings” (Donovan, Smolkin, & Lomax, 2000, p. 329). However, in

general there is also an undeniable emphasis on matching books and readers, such that it has even been described as a “leveling mania” (Dzaldov & Peterson, 2005; Szymusiak & Sibberson, 2001).

The other side of this focus on text difficulty is that students are also frequently warned away from texts that may be too difficult for them, with the assumption that these texts are unhelpful and even harmful. Indeed, perhaps the most common term for too-difficult texts is “frustration-level,” a term that originated from the writings of Emmett Betts (1946), whose framework of reading levels will be discussed in more detail in Chapter 2 of this dissertation. These frustration-level texts were so labeled because they were believed to prompt frustration and anxiety in the students who read them. Even today, scholars argue that time spent with too-difficult texts can reinforce bad reading habits and decrease motivation to read, while sacrificing valuable time that could have otherwise been spent on more enriching reading experiences (e.g., Fountas & Pinnell, 1996; Allington, 2006; Afflerbach, 2007).

However, there are also those who argue that difficult texts can be valuable for students for a number of reasons. Hunt (1970) eloquently described the benefits of pairing “the high-interest book and the low-powered reader,” contending that interest in a topic can enable a reader to learn from and enjoy a text that would otherwise be considered too difficult (p. 147). Similarly, Worthy and Sailors (2001) have argued that methods of determining what texts are appropriate for specific students should be expanded beyond simple readability to include factors such as motivation, interest, and prior knowledge. In addition, Brown (2000) mentions that difficult texts are sometimes the only avenues to certain content, stating that “accessibility has an inverse relationship

with complexity. That is, the *most accessible* texts also are those with the *least complex* content” (p. 295, emphasis in original). In other words, the simplification that may make a text more readable may lead to a loss of complexity and an alteration of the content itself, such that the content of a difficult text may be fundamentally different from the content of a simplified text on a similar topic (Brown, 2000). These comments suggest that situational features such as interest, choice, and purpose may influence whether or not students experience frustration when they read difficult texts.

The ongoing arguments about how strongly (if at all) to emphasize or require an ability-readability match between readers and texts are complicated by the fact that there is little research to substantiate the claims on either side of the debate. As the review of literature in the following chapter illustrates, significant questions remain about the impact of text difficulty on students’ reading experiences, especially in the case of texts that students read independently.

The purpose of this study was to fill some of the gaps in the existing research by exploring and describing the experiences of a sample of second grade students who chose to read texts that would typically be considered too difficult to be read independently. The study was situated in the context of classroom-based independent reading time, which is a period of the day during which students choose freely from texts in the classroom library and read silently to themselves. The study was designed to directly address the relationships between text difficulty, reading comprehension, and reading motivation. It addressed the following set of research questions:

1. Do the students who choose texts at their frustration level fit into certain profiles based on gender, motivation to read, or reading ability?

2. What reasons or purposes do students give for choosing frustration-level texts for independent reading?
3. What, if anything, do students understand from reading self-selected, frustration-level texts (as compared to independent-level texts) independently?
4. What are students' perceptions of the difficulty of their chosen, frustration-level texts?
5. What are students' affective experiences with reading these difficult texts independently? In other words, are self-selected, frustration-level texts actually "frustrating," or are students able to enjoy them?

Five second grade classrooms were selected for participation in the study, and all students in these classrooms completed whole-class assessments of reading ability and motivation to read. During their regular independent reading time, students kept written reading logs, in which they recorded information about the texts they chose, their reasons for choosing them, their perceptions of the texts' difficulty, and their enjoyment of the texts. Information from the reading ability test and from the reading logs was used to identify occasions in which students appeared to have chosen texts that may have been difficult for them. For each of these occasions, students participated in a series of three one-on-one reading assessments: an oral retelling of the selected text, an assessment of oral reading accuracy on a portion of the selected text, and a set of text-based comprehension questions. Following these three assessments, students answered questions about their enjoyment of their selected texts and their perceptions of the texts' difficulty for them. For every occasion in which a student was assessed reading a frustration-level text, a paired occasion was identified in which the same student chose to

read an easier text. In some cases, in which no easier texts were identified from the list of texts students had chosen to read, researchers provided students with a smaller set of easier texts from which to choose. As a point of comparison, the same assessments and interview protocol were used to gather information about students' experiences with the easier texts. To address the five research questions listed above, this study employed a mixed-methods design, incorporating data gathered from multiple sources and using both quantitative and qualitative analyses. Quantitative analyses included t-tests and correlations, while qualitative work included coding of students' reading log entries and interview responses.

This study contributes to a better understanding of relationships between text difficulty, reading comprehension, and reading motivation in the context of self-selected, independent reading. Information about these relationships is important for primary grade teachers, for whom the issue of matching readers with texts that will facilitate reading growth is a central concern. In particular, given the ubiquity of Betts's (1946) framework for making reader-text matches, it is important to understand how useful it actually is for predicting the quality of students' reading experiences, in terms of both comprehension and motivation. The issue is also important because students frequently encounter texts that would typically be deemed too difficult for them, either as assigned, instructional texts or as self-selected, independent reading material. By looking more closely at students' experiences with difficult texts, it is possible to offer some new insights and related recommendations for the role that text difficulty should play in matching elementary readers to texts. In particular, the findings from this study provide important information about any potential value that difficult texts might hold for the readers who

choose them, as well as ways that teachers might support and scaffold students' reading experiences with them.

Chapter 2 of this dissertation describes the relevant literature pertaining to issues of text difficulty, reading motivation, and reading comprehension. It also describes the broader theoretical works and ideas that have informed the study design and data analysis. Chapter 3 describes the study methods in detail, including administration and scoring procedures for each of the individual data sources used in the study. Chapter 4 offers the results of the analyses related to each of the five research questions listed above. Chapter 5 provides an in-depth discussion of the study's findings, including limitations and possible directions for future research.



## CHAPTER 2: LITERATURE REVIEW

### Overview

The issue of text readability is not new; it has been a consideration of teachers, researchers, and publishers for many years. Starting in the early 1920s, researchers began to address questions of text readability by finding new ways to predict how difficult a text might be for its readers, first based on vocabulary analysis and later based on additional factors largely related to syntactic complexity (e.g., Dolch, 1928; Thorndike, 1921; see Chall & Conard, 1991, for a review). The presumed importance of this early research on text readability was based on the central idea that text readability matters for reading and learning. But just as texts vary in their readability, readers vary according what they bring to the reading experience – factors including knowledge, attitudes, and skills. Thus, text readability came to be viewed as only meaningful in terms of its relationship to individual readers. As a result, early research in text readability was followed by research that endeavored to calibrate and align established levels of text readability with measures of individual reading ability, effectively moving from analyses of text readability to discussions of suitable difficulty relative to individual readers (Chall & Conard, 1991).

Before going further, it is important to note the differences between the terms “readability” and “difficulty.” For the purposes of this dissertation, I use the term *text readability* to refer to a text-specific characteristic; for example, how readable is a given text for readers in general, based on text factors such as sentence length, semantic complexity, and so on? In contrast, I use the term *text difficulty* to refer to the relationship between a specific text and a specific reader. In other words, the difficulty of a text is a function of the text’s readability in relation to the reader’s reading ability, in addition to

other factors such as interest, motivation, and background knowledge. Difficulty is also situated within a larger context, as the difficulty of a particular text for a particular student may also depend on contextual factors, such as the presence or absence of material, interpersonal, or interactional scaffolds.

This distinction between readability and difficulty arises from the research base surrounding reading comprehension, which has shown that reader factors such as interest (e.g., Asher & Markell, 1974; Asher, Hymel, & Wigfield, 1978; Bernstein, 1955; Lin, Zabrocky, & Moore, 1997), prior knowledge (e.g., Baldwin, Peleg-Bruckner, & McClintock, 1985; Kintsch & Franzke, 1995; Voss & Silfies, 1996; Wolfe & Mienko, 2007), and motivation (e.g., Baker & Wigfield, 1999; Guthrie et al., 2007; Wigfield et al., 2008) can affect an individual reader's comprehension of a text. Rather than a direct connection between text readability and comprehension, other variables mediate the relationship and determine the actual difficulty of a particular text for a particular student in a particular context. The theoretical importance of this distinction is that difficulty is relational rather than fixed – it varies based on relationships between readers, texts, and tasks. It is important to note, however, that my personal observations and experiences suggest that this distinction tends to be overlooked in actual classroom practice. Despite the theoretical differences between readability and difficulty, it seems that many reader-text matches are based on simple determinations of student reading level (based on word identification ability) and measures of text readability (based largely on word and sentence length).

### *Betts' Reading Level Framework*

One of the most influential frameworks for determining the appropriateness of particular texts for particular readers is the system of reading levels originated by Emmett Betts. In his 1946 book *Foundations of Reading Instruction*, Betts laid out a framework that describes four levels of text difficulty: 1) the basal, or independent, level, which is “the highest reading level at which the individual can read with full understanding and freedom from mechanical difficulties”; 2) the instructional level, which is “the highest reading level at which systematic instruction can be initiated”; 3) the frustration level, which is the level at which a reader is “thwarted or baffled by the language (i.e., vocabulary, structure, sentence length) of the reading material”; and 4) the probable capacity level, which is “the highest reading level at which the individual can comprehend (i.e., deal adequately with the facts by means of oral language) material read to him” (p. 438-439).

While Betts' probable capacity level is based on listening comprehension, he describes the other three reading levels as being determined by assessing a student's oral reading accuracy and comprehension of texts written at varying readability levels. The use of decoding accuracy as a primary criterion for determining reading levels – as seen in Betts' seminal work (1946) and in a number of contemporary, published reading assessments (e.g., Johns, 1997; Stieglitz, 2002; Woods & Moe, 2007) – initially derives from Kilgallon's (1942) finding that most of the reading difficulties experienced by his sample of fourth grade readers were word-perception errors. According to Betts (1946), students are reading at their independent level when they demonstrate at least 99% accuracy in their oral reading and 90% or higher comprehension. The standards for

instructional-level are slightly lower, at between 95 and 99% oral reading accuracy and between 75 and 89% comprehension. And a student's frustration level can be identified when either his oral reading accuracy has dropped to 90% or less, or his comprehension score is 50% or lower (Betts, 1946). It is important to note that although Betts defined the independent and instructional levels by using *both* word recognition and comprehension criteria, the frustration level can be established if *either* of the two measures is at or below the levels he describes.

In addition to this set of criteria related to reading performance, Betts (1946) goes on to list behavioral indicators that he considered to be characteristic of students' silent reading at each of the levels. For example, he explains that oral reading of independent- and instructional-level texts should be characterized by a lack of vocalization and finger pointing, while frustration-level texts may lead students to omit and reverse words (Betts, 1946). Because the proposed study focuses largely on frustration-level texts, it seems important to describe this level – as Betts originally conceived it – in a bit more detail.

In describing students' behaviors during reading of frustration level texts, Betts states, "As the typical pupil becomes increasingly frustrated, he may exhibit tension, movements of the body, hands, and feet, he may frown and squint, and he may exhibit other types of emotional behavior characteristic of a frustrated individual" (p. 445). In a later, in-depth description of frustration-level reading, Betts (1946) lists ten criteria for estimating the frustration level, including physical behaviors such as irregular breathing, frowning, erratic body movements, whispering, and crying. His description of frustration-level reading from a pupil's view labels the frustration level the "troublesome step" and includes the following description:

I don't understand half of what I read and I feel worried and unhappy. Sometimes I can't stand still as I read and I'd like to point with my fingers or read with my lips. Often I can't pay attention. I should not read on this step often for my reading will not improve. (Betts, 1946, p. 448)

Taken together, these descriptions clearly show the assumed link between lack of oral reading accuracy, limited reading comprehension, and frustration as an emotional response.

### *Influence of Betts' Framework*

Betts' framework is significant because of the ways it has influenced and continues to influence reading assessment and instruction. The framework remains widely used today, as many teachers use Betts' guidelines to help them determine students' reading levels and match students to appropriate texts. Perhaps the most obvious way that Betts' ideas have worked their way into the daily practice of elementary classroom teachers is through their role in a variety of reading assessments, particularly those known as Informal Reading Inventories (IRIs). These IRIs, which Betts is frequently credited with pioneering, typically require students to read from a series of word lists and written passages, which are presented in order of increasing difficulty. The administrator assesses the student's ability to read words accurately in the lists and in the written passages and to comprehend the written passages. The results of the IRI can be used for diagnostic purposes, as teachers analyze the types of decoding and comprehension errors that a student makes, but they can also be used to match a student with appropriate reading materials by determining his or her independent, instructional, and frustration reading levels. IRIs are ubiquitous in elementary classrooms and have had

a heavy influence on how Betts' framework has been put into practice.

In addition to his influence on reading assessment, Betts has influenced reading instruction and policy through his recommendations for appropriate uses of texts at varying levels of difficulty. In his 1957 discussion of frustration-level texts, Betts describes the frustration level as the level "to be avoided...the material is too difficult and frustrates the reader" (p. 448). As described earlier, numerous scholars and authors have echoed Betts' advice and warned teachers about the assumed pitfalls of having readers engage with texts that present significant challenges in terms of word decoding and comprehension (e.g., Lipson & Wixson, 2003; Mariotti & Homan, 2005; Robb, 2000; Stahl, 2004; Tompkins, 2006). Despite these recommendations, however, students still encounter frustration-level texts, both as assigned reading and as independent reading choices.

### *Critiques and Revisions*

In the years since Betts first proposed his reading level framework, it has come under criticism from a number of directions, including the methods used for determining the levels, the criteria that define the levels, and the applications of the framework to classroom practice. For example, Hunt (1970) warned teachers against an over-reliance on IRIs, questioning the very idea that assessments of student reading performance should focus on errors rather than successes. He also offered a critique of IRI scoring methods, arguing that the practice of using running records to calculate a percentage of oral reading errors assumes incorrectly that all reading errors are equally harmful. Regarding assessment procedures, Jongsma & Jongsma (1981) compared a number of available IRIs and found that they sometimes used slightly different administration

procedures and scoring rules. For example, Betts recommended that students read the passages silently before having them read them aloud, but not all IRIs follow this model. In addition, there has been some debate about what should count as errors, particularly in the case of repetitions (e.g., Davis, 1975; Jongsma & Jongsma, 1981). These differences in administration and scoring logically lead to variance in results.

In addition to criticisms of IRI methods and scoring, Betts' criteria have also been questioned. Hunt (1970) doubted the claim that good reading must necessarily be error-free or nearly error-free, since the purpose of reading is to get ideas and meaning from text. He felt that the 99-100% accuracy rate was unreasonable, especially for young readers. He also questioned the 50% comprehension cutoff, commenting that not all ideas in a text are equally important for students to learn. Powell (1970; see also Powell and Dunkeld, 1971) offered an additional challenge to Betts' numbers, arguing that there were developmental differences in word reading accuracy. He felt that a 95% threshold of word reading was too high for primary grade readers, particularly when reading instructional texts, and suggested a revised framework with different leveling criteria depending on the age of the reader. Powell and Dunkeld (1971) sought a more accurate definition of the instructional level in particular, commenting, "The instructional level designated by the IRI is an unvalidated construct" (p. 637).

And finally, the use of Betts' levels for matching readers to texts has also been contested by a number of scholars, largely on the grounds that, while placement levels based on word reading accuracy and comprehension proficiency may serve as helpful guidelines in making text matching decisions, they are not the only criteria that teachers and students should consider (e.g., Dzaldov & Peterson, 2005; Worthy & Sailors. 2001).

Specifically, these scholars suggest consideration of other factors, including interest, prior knowledge, cultural background, and enjoyment (e.g., Rhodes & Dudley-Marling, 1996; Worthy & Sailors, 2001). Given this set of challenges to both the common methods of determining text difficulty and the popular application of Betts' reading levels to instructional situations, the present study aims to find out more about students' experiences with texts of varying levels of difficulty. Additional information about the students who choose difficult texts, their comprehension of those texts, and their enjoyment of the texts may contribute valuable findings to the current debate over reader-text matches.

### Research Base

The following sections describe the existing research base related to each of the five research questions addressed by this study.

*Question 1: Do the students who choose texts at their frustration level fit into certain profiles based on gender, motivation to read, or reading ability?*

Students' text choices have been a topic of study for many years. While much research in this area has focused on genre preferences, a number of studies have also addressed issues of text difficulty. Of particular relevance to this review are studies that have examined relationships between student variables (such as gender, age, and reading ability) and reading choices relative to text difficulty.

A common finding in this area has been that many students choose books that would typically be considered too difficult to be read independently (e.g., Anderson, Higgins, & Wurster, 1985; Kragler, 2000; Mork, 1973). Another common finding is of relationships between student reading ability and the tendency to choose difficult texts.



For example, in his sample of sixth grade students, Olson (1984) identified a statistically significant tendency for low readers to choose above-level texts and a non-significant tendency for high readers to choose below-level texts. This finding mirrors Fresch's (1995) finding that high-ability readers were somewhat less likely than average and below-average readers to choose books far above their reading levels. Similarly, Kragler (2000) studied fourth grade boys and found that low readers often chose frustration-level texts while high readers frequently chose books that were too easy. Donovan, Smolkin, and Lomax (2000) found that the tendency to choose difficult texts was most pronounced among low-ability readers, although all of the first graders in their study chose difficult texts at least occasionally.

While these studies appear to document a trend of above-level choices for below-level readers, Lysaker's (1997) work offers conflicting results. In her sample of first grade readers, the less-skilled readers chose easy books that required little, if any, actual word reading, while the more-skilled readers chose a wider range of more challenging texts. Other studies have found that, rather than consistently selecting texts in a certain range of difficulty, young readers of all levels of reading ability actually cycle back and forth between easier and more difficult texts (Timion, 1992; Fresch, 1995).

Other studies have looked beyond reading ability to explore the influence of additional variables on the reading level of students' text choices. For example, Hiebert and colleagues (Hiebert, Mervar, & Person, 1990) considered instructional context by comparing the book selections of second graders from five literature-based classrooms and five textbook-based classrooms. In their sample, the students most likely to choose books above their grade level in terms of readability were the less-skilled readers in

textbook-based classrooms. Olson (1984) investigated the role of gender in text selection, finding no significant relationship between gender and the readability levels of text choices. It is important to point out that these two studies both dealt with the issue of text difficulty by using readability formulae to determine grade levels; the findings therefore reflect general levels of text readability rather than actual text difficulty for the students who chose the texts.

One potentially important, yet understudied, factor in considering the question of which students choose difficult texts is motivation to read. Two aspects of motivation to read that may be particularly relevant to situations involving student choice are individuals' perceptions of themselves as readers and the value they place on reading. In the particular case of text difficulty, students who choose difficult texts may be those who have higher estimations of their reading ability. Additionally, the degree to which students value reading may influence their choices of more or less difficult texts for independent reading. The present study's inclusion of a quantitative measure of motivation to read – assessing students' perceptions of themselves as readers and the value they place on reading – will help explore the possible roles that these constructs play in student text choices.

*Question 2: What reasons or purposes do students give for choosing frustration-level texts for independent reading?*

Perhaps the most consistent finding from the research on student text choices is that students of both genders, of all levels of reading ability, and with varying classroom reading experiences all choose difficult texts at some point. The logical follow-up question to this phenomenon relates to why they make these choices. Fielding and Roller

(1992) offer three possible explanations, suggesting that children either a) don't know how to find appropriate-level books; b) don't have access to appropriate-level books; or c) don't want to read appropriate-level books. Moss and McDonald (2004) suggest that struggling readers might choose difficult texts that offer high levels of non-print support because these texts contain desired information and can be "read" without actually decoding the print. Hunt (1970) similarly ascribes the choice of difficult texts to students' desire to pursue their interests, arguing that "the reader who finds a really good book for him, the book that has ideas he truly wants to learn about, frequently will outdo his own instructional level of performance" (p. 148). Donovan and colleagues (2000) also offer a set of possible explanations for their finding that children of all levels of reading ability apparently chose difficult texts. They suggest that students may have made these choices because of: 1) pure chance, based on the readability levels of available texts; 2) familiarity with the challenging texts, which were often books that had been introduced through teacher read-alouds (which may have effectively rendered them less difficult for the young readers); 3) interest in a specific topic; and 4) a desire to "transform the texts" by reading them in small groups and acting them out as a social activity (p. 328). Another possibility that they mention only briefly is that some students "wanted to read what the 'good readers' were reading" (p. 326). However, because this study only included data on students' choices and not on the reasons behind their choices, any explanations on this front are conjecture (Donovan et al., 2000).

In summary, the question of why students choose difficult texts remains largely unanswered. Although a number of appealing hypotheses have been put forth, there is currently little research evidence to confirm or refute them. One explanation that does

have some empirical support is the notion that students sometimes choose difficult books for social reasons. They may choose difficult texts either because those texts are popular with peers or because they are trying to establish an individual identity as a “good reader.” For example, Fresch (1995) found that first grade students wanted to read books that were popular with their classmates. Bass (2006) found that some first graders viewed chapter books as status symbols and used them to help define themselves as good readers. In a verbal protocol study of students’ text choices during visits to the school library, Halladay (2006) found this desire to be seen as a “chapter book reader” to be true for some second graders as well. This present study aimed to build on the small base of empirical evidence regarding reasons for choosing difficult texts and to test some of the contentions that have been put forth by other researchers in this area.

*Question 3: What, if anything, do students understand from reading self-selected, frustration-level texts (as compared to independent-level texts) independently?*

As discussed earlier, many have touted the importance of ability-readability matches and have warned against the dangers of having students engage with frustration-level texts, based in part on the assumption that students’ ability to derive meaning, gain content knowledge, and develop reading skills is hindered by increasing levels of text difficulty. However, despite the seeming authority with which these recommendations are often made, there is actually little research on the comparative benefits of reading independent, instructional, and frustration level texts.

In summarizing the literature in this area, it is important to be clear whether the texts under investigation were used for instruction or for recreational reading, and whether they were assigned to students or chosen by students. Regarding assigned,

instructional texts, several studies have sought to determine the impact that text difficulty plays in students' development of reading fluency, use of comprehension strategies, text comprehension, and vocabulary acquisition. The following paragraphs summarize the extant literature in this area.

In terms of reading fluency, simple texts may be more effective than more difficult texts in improving fluency for young readers. For example, Hiebert (2005) used an experimental design to measure second graders' fluency gains in three conditions: 1) a treatment using repeated readings of content-area texts with few rare, multisyllabic words; 2) a treatment using repeated readings of literature selections with more rare, multisyllabic words; and 3) a control group who received their standard literature-based instruction. Of these three conditions, the content-area group made significantly larger gains in fluency than did students in either of the other two conditions. These gains in oral reading rate and accuracy were most pronounced for students who entered the study with below-average fluency scores. However, although the two treatment conditions both outperformed the control group in comprehension gains, there were negligible differences between the two treatment groups on this measure.

In another fluency-related study, O'Connor and colleagues (O'Connor, Bell, Harty, Larking, Sackor, & Zigmond, 2002) also used three conditions – two treatment and one control – to explore the role of text difficulty in reading skill gains for fifth grade readers involved in a one-on-one tutoring program. The two treatment conditions used an identical tutoring approach but differed in their use of either a) texts matched to students' reading level (Reading Level Match – RLM) or b) texts matched to students' grade level (Classroom Match). They found that both treatment conditions outperformed the control

group on oral reading fluency, and the RLM group performed better than the CM group. Another interesting finding was that students who began with lower fluency did better in the RLM condition; for those who began with higher fluency, the RLM and CM treatments were equally effective. One conclusion was that RLM texts are good for building fluency in low fluency readers, possibly because of the high redundancy of words and frequent repetitions. Taken together, these two fluency-related studies suggest that ability-readability matches may benefit struggling readers, at least in terms of oral reading fluency in instructional situations.

Text difficulty may also influence readers' use of comprehension strategies. For example, Kletzien (1991) found that high school readers' use of comprehension strategies differed in relation to the difficulty of the texts they read. In this study, the texts in question were assigned passages from social studies textbooks. Kletzien found that good readers and poor readers used the same type and number of strategies on easy passages, but poor readers used fewer comprehension strategies on difficult passages. In contrast, as passage difficulty increased, good comprehenders actually used a wider variety of strategies and used them more often than their less-skilled peers.

In another comprehension-related study, White and Jordan (1987) found that adult learners in a vocational technology program were able to read assigned, field-specific materials even when their reading level scores suggested that they would have difficulty with the materials. They concluded that background knowledge plays a big enough role in comprehension that readers can understand texts that would otherwise be inaccessible. This study is similar in its findings to a number of other studies that have addressed

relationships between prior knowledge and reading comprehension, except that it also considers text readability as a factor.

One final example of research into the relationship between text difficulty and reading gains for instructional texts is a study by Eldredge (1990), who used two conditions in his work with struggling third grade readers over an 8-week treatment period. One group was assigned high-interest passages at frustration level and was provided with instructional scaffolds during reading. The other group was allowed to choose from a large set of books – with explicit instructions to choose texts they could read independently – and was given no additional assistance. The assisted group outperformed the unassisted group on pretest-posttest gains in vocabulary and reading comprehension, and Eldredge concluded that third grade readers were most successful in improving their reading skills when they were challenged by the texts they read and assisted during their reading of them. Based on these findings, Eldredge recommended both difficult texts and scaffolded instruction, but he did not recommend having students read difficult texts independently. However, one problem with this study is that the comparison involves too many variables to be able to isolate any single cause of the outcomes, because the two groups differed on number of texts read, text difficulty, presence or absence of scaffolding, and chosen or assigned books. Additionally, because the study only included fictional, narrative texts, the findings cannot be generalized to reading experiences with texts from other genres.

Regarding texts read independently, numerous studies have examined the effects of reading volume on reading achievement, with mixed – and often hotly contested – results (e.g., Anderson, Wilson, & Fielding, 1998; Davis, 1988; Guthrie, Wigfield,

Metsala, & Cox, 2004; Kamil, 2007; Krashen, 2005; Nagy & Scott, 2004; Stanovich, 1986; Taylor, Frye, & Maruyama, 1990). Because few studies of reading volume and reading achievement have used text difficulty as a variable, the differential effects of engaging independently with easy and difficult texts are also uncertain. Davis (1988) found that medium-ability eighth graders showed larger reading gains from independent reading than did their high-ability peers, and he suggested that this may have been because the texts they were reading were closer to an instructional level for them. In contrast, he suggests that the high readers may have been reading books that were too easy for them to make similar gains. However, in the absence of specific information about student reading level and text readability, these notions remain untested hypotheses.

Carver and Leibert's (1995) investigation of text difficulty, independent reading, and reading achievement is perhaps the best-known study in this area. The authors undertook this study as a way to find out if there is such a thing as a reading bootstrap; in other words, "can they simply read and lift themselves up by their bootstraps to a higher level of ability that probably involves more difficult vocabulary (less frequent words) and more complex knowledge structures" (p. 26)? The authors theorized that bootstrapping would be impossible for easy reading because simple texts offer no new information and demand no conceptually complex thinking. But they argued that bootstrapping should also be impossible (or at least problematic) for difficult texts, because "there is no known learning mechanism that would allow this bootstrapping to occur" (p. 31). Working with a sample of third, fourth, and fifth grade students in a six-week summer school program, they assigned each student to one of two conditions: easy reading or matched reading.



The first group was allowed to choose any books that were at or below their instructional reading level, and the second group was allowed to choose books that were at or slightly above their instructional reading level. At the end of the thirty-day period, during which students completed between 15 and 30 hours of reading, students were assessed on a battery of reading skills. Carver and Leibert found no gains for either group. However, the authors are careful to point out several possible explanations for the lack of change in reading skill, including the short duration of the study and some discrepancies in reading leveling systems that resulted in negligible differences between the actual readability of the books in the two groups.

One difficulty with research on this topic is that, when using Betts' (1946) set of reading levels to define text difficulty, any discussion of comprehension and text difficulty is complicated and confounded by the fact that comprehension itself is a central component of the criteria for establishing the levels. So the very question of what students understand from frustration-level texts seems like circular logic: what do students understand from texts that they, by definition, do not understand? However, there are several reasons why a question like this is still sound and worthy of investigation. First, many applications of Betts' reading levels, including Betts' original criteria, call for *either* 90% or less word reading accuracy *or* 50% or less reading comprehension, but not both. Because of this, it is technically possible for a student to read a frustration-level text in terms of oral reading accuracy with better than frustration-level comprehension. In fact, a recent study by Riddle Buly and Valencia (2002), found that a number of students fit this profile, with remarkably good comprehension despite significant problems with word decoding. Second, limited understanding is not the same

as no understanding. So the real question in this case is: if a student is assessed as having less than 50 percent comprehension for a given text, what does the up to 50 percent comprehension contain? Hunt (1970) argues this point eloquently, saying that a few ideas important to the reader can be more important than all the ideas important to the teacher. And third, on a related note, the very notion of assigning a percentage score to represent the amount that a reader comprehends from reading a text seems problematic, suggesting a need for a more qualitative look at students' understanding of the texts they choose to read.

*Question 4: What are students' perceptions of the difficulty of their chosen, frustration-level texts?*

Numerous studies have addressed students' perceptions of their own reading ability, finding that older readers tend to be more realistic and less optimistic in their self-assessments than younger readers are (see Pressley, 2006 for a summary of this research). However, only a small number of studies have looked directly at students' perceptions of text readability or difficulty. O'Hear and Ramsey (1990) worked with college students to find out whether there was a match between student perception of reading ease and actual readability of three different college composition texts. By comparing student ratings on a Likert scale with analysis from five commonly used readability formulas, they found large discrepancies between student perceptions and formula estimations. In a subsequent study, Ramsey (1994) again examined college students' perceptions of readability of upper-level college composition texts and found that their perceptions were not aligned with readability determinations based on Flesch (1948) and Fry (1977) readability formulas. The only study on this topic that involved elementary students is Fleming's

(1967) investigation of fifth graders' perceptions of text difficulty. Using 32 passages at different readability levels and on different topics, Fleming found little consistency in the students' ratings of materials as easy or hard. For all of these studies, it is important to note that the identified discrepancies between measured readability and perceived difficulty may derive from several sources: inaccuracies in student perceptions, inaccuracies of readability formulas, or differences between general readability and relative difficulty for a particular reader.

Although there is little research on this topic, it has some potentially important practical and theoretical implications. Knowing whether or not students are able to accurately assess a text's difficulty for them would be useful information for teachers and researchers alike.

*Question 5: What are students' affective experiences with reading these difficult texts independently? In other words, are self-selected, frustration-level texts actually "frustrating," or are students able to enjoy them?*

Since Betts first coined the term "frustration-level," probably the strongest argument against engaging students with difficult texts is that they will cause some degree of mental and physiological stress to the readers who must struggle to read them. However, despite the prevalence of this argument, there is actually very little evidence to support the connections between levels of difficulty – as measured by rates of oral reading errors and comprehension – and frustration. When Betts created the original framework in the 1940s, he established categories and assigned characteristics to each of them without testing their validity. In the more than 60 years that have passed since then, few studies have attempted to directly validate the assumed, affective aspects of the

frustration reading level. In fact, most of the literature in this area comes from a small series of studies in the 1970s.

Ekwall, Solis, and Solis (1973) used a polygraph to measure children's physiological responses to the potentially stress-inducing situations of reading passages at progressively higher levels of readability. Working with a sample of 62 third, fourth, and fifth graders, they compared the students' rate of comprehension for each passage with the polygraph measures of blood pressure, heartbeat, breathing rate, and perspiration that would indicate an emotional reaction. They found that most children were not actually physiologically frustrated at the 50 percent comprehension level identified by Betts (1946). This finding was particularly true of poor readers. However, some of the higher ability readers in their sample showed signs of frustration even when they were reading at comprehension levels higher than 50 percent, possibly because of their high self-concepts.

In a related study, Davis (1975) also used the polygraph to measure frustration, but he focused both on comprehension and on oral reading errors. His goal was to test the accuracy of Betts' frustration level benchmarks of 50 percent or less comprehension and 90 percent or less oral reading accuracy. Using Spache's (1963) Diagnostic Reading Scales, Davis measured oral reading accuracy on leveled passages and comprehension of passage-specific questions. He found that students' polygraph indications of frustration were aligned with a mean *comprehension error* rate of 58.39% (SD = 21.64%), suggesting a corresponding *comprehension* rate of 41.61% that is quite a bit lower than Betts' 50% criterion. Regarding oral reading accuracy, Davis found that when repetitions were counted as errors, Betts' 10 percent error rate corresponded fairly well with

polygraph indications of frustration. However, when repetitions were not counted as errors, polygraph readings generally indicated frustration before students reached the 10 percent error rate.

The third article in this series of polygraph studies (Davis & Ekwall, 1976) used the same dataset as the previous one, but it included personality type as a student variable to determine the nature of the apparent individual differences in frustration level. The authors reported that 87% of their sample of third, fourth, and fifth grade students reached polygraph frustration at a level of about 6-9 percent oral reading errors. The group of students who did not get frustrated at that level – and in many cases until nearly twice that level – did not cluster along the individual variables of gender, age, or intellectual ability, but they did share a certain personality profile, described as a “restricted” mode of perception (p. 452). According to the authors, individuals in this category are characterized by the tendency to persist with initial perceptions, a lack of imagination, and a failure to consider multiple, possible solutions to a problem: “For him there are no ambiguities. If the first perceived solution does not work, the fault is with the problem, not with the rigid adherence to the first option for a solution” (Davis & Ekwall, 1976, p. 452). The authors concluded that, “for most children, reading passages for instructional purposes must be no more difficult than to allow for about 5% oral reading errors” (p. 453). However, it is important to note that this study only uses assigned passages, so the results are not directly generalizable to reading in choice settings.

As this last statement suggests, one possible limitation of the Betts framework that merits further investigation is that the reading levels are meant to apply to reader-text matches writ large, without much consideration of context. Although the descriptions of

the levels do mention the amount of external support required for successful reading, the overall framework foregrounds the relationship between a reader's ability and a text's readability, overlooking the possible influences of factors such as choice and purpose. Regarding frustration-level texts in particular, it seems possible that these factors help determine whether or not a reading experience will be frustrating for a given reader, even beyond the role of the text's readability. This point becomes critically important when we consider the research base that supports the role of choice in student motivation and performance. Hunt (1970) emphasized the importance of choice when he described situations in which a reader may be able to transcend the frustration level:

When the classroom atmosphere encourages self-selection, usual reading level performances become less meaningful. This author has watched many readers spend many rewarding moments with material which by any standard inventory would be classified as too difficult. (p. 148)

He goes further to argue that, when a student "has chosen the material to read because of personal interest, he can break many of the barriers" (p. 148). On a related note, Dzaldov and Peterson (2005) have argued that restricting choices to leveled texts may actually "dampen students' motivation to read" (p. 223). And Donovan and her colleagues (2000) suggest that "access to difficult informational texts may have provided at least some children the opportunity to develop, or enhance, intrinsic motivation to read that is so important to successful reading achievement" (p. 329-330).

### *Summary*

As this review of the literature illustrates, some important questions remain regarding the role of text difficulty in the reading process, particularly in the case of self-

selected, independent reading. Further research in this area is important, given the frequency with which teachers appear, based on my observations and experiences, to apply Betts' framework to make instructional decisions. And it seems especially important in light of the fact that the evidentiary base for the framework does not seem to warrant its current ubiquitous use or the strict adherence with which it is usually applied. In other words, although Betts' framework and its subsequent revisions and applications have created distinct categories of reading level and defined them with specific percentages of reading accuracy, knowledge about students' actual reading at each of these levels is considerably fuzzier. This study was intended to strengthen the research base related to reading and text difficulty by finding out more about students' reasons for and experiences with reading texts of varying levels of difficulty, especially frustration-level texts.

### Theoretical Framework

Having reviewed the relevant research literature, I now describe some of the theoretical ideas that are important to this study. These theories informed the design of the study, and they are also helpful in defining a set of central constructs. Specifically, I outline some of the major theoretical works that have influenced my thinking in relation to each of the three main constructs under consideration in this study: reading comprehension, text difficulty, and motivation to read.

#### *Reading Comprehension*

For a term that seems like it should be self-explanatory, "reading comprehension" has been theorized in a variety of different ways; there are numerous models of reading comprehension processes as well as ongoing conversations about what actually counts as

comprehension and how it can be measured. As one scholar put it, “Understanding and comprehension are everyday terms, useful, but imprecise” (Kintsch, 2004, p. 1270). Different theoretical models of reading comprehension put varying degrees of emphasis on a reader’s cognitive and perceptual processes (e.g., LaBerge and Samuels, 1974; Gough, 1972; Rumelhart, 1994); text structure (e.g., Kintsch, 2004); reader-text interactions (e.g., Rosenblatt, 1978; 1994); readers’ affective factors, such as attitude, motivation, and interest (e.g. Mathewson, 1976, 1985, 1994); and social and instructional contexts (e.g., Ruddell and Unrau, 2004). For any study that focuses on reading comprehension, then, it is important to be clear about which model or models of comprehension inform the research and what implications they have for the study’s design, implementation, and analysis. This study draws heavily on one particular model of comprehension, developed by the RAND Reading Study Group (2002).

The RAND Reading Study Group was a 14-member panel funded by the United States Department of Education’s Office of Educational Research and Improvement (OERI). The group’s task was to review the existing research literature related to reading comprehension and to propose a set of “strategic guidelines for a long-term research and development program supporting the improvement of reading comprehension” (RAND, 2002, p. iii). The second chapter of the report offers a formal, detailed definition of reading comprehension, which serves as the foundation for their proposed research agenda. The report defines reading comprehension as “the process of simultaneously extracting and constructing meaning through interaction and involvement with written language” (p. 11). By this definition, the text itself is important but insufficient for determining reading comprehension. The report goes on to describe a model of reading



comprehension that includes three main elements: “the reader, the text, and the activity or purpose for reading” (RAND, 2002, p. xiii). A fourth element of the RAND model is the sociocultural context in which readers engage with texts and make meaning from them. The three other components – reader, text, and activity – are said to exist and interrelate within this broader, sociocultural context “that shapes and is shaped by the reader and that interacts with each of the elements iteratively throughout the process of reading” (2002, p. xiii).

The influence of the RAND model on this study is evident in several important ways. First, the depiction of reading comprehension as being comprised of three interrelated components – reader, text, and activity – is reflected in the study’s multiple focuses on student characteristics, text characteristics, and purposes and instructional contexts. Second, each of the three components is considered as a complex construct rather than as a one-dimensional item. For example, the term “reader” is understood to include not just generalized reading ability, but also the interest, motivation, knowledge, skills, and experiences that individual readers bring to bear on the act of reading specific texts. Second, the model presupposes that meaning resides not solely in the text itself but in more complex interactions between readers and texts, relative to the varied activities in which reading occurs. And finally, just as the model prioritizes the sociocultural context that surrounds the reading process, the study is grounded in the daily life of individual classrooms, where individual students’ interactions with texts are influenced by aspects of the surrounding physical, social, and instructional environment.

## *Text Difficulty*

As mentioned in the earlier literature review, Betts' (1946) framework of independent, instructional, and frustration reading levels serves as this study's primary theoretical foundation for the construct of text difficulty. Because this framework has already been discussed in some detail, it will not be addressed further in this section, which first focuses on the broader topic of task difficulty and then moves to issues related specifically to reading.

Much of the research related to task difficulty has been in the field of psychology and has focused on finding out what level of challenge – or task demand – individuals are willing to take on (e.g., Atkinson & Feather, 1966; Sagie, 1993; Spence & Helmreich, 1983; Weiner, 1972). This body of literature will be discussed further in the upcoming section dealing with motivation to read. Other research has focused on relationships between task difficulty and adverse outcomes such as problem behavior (e.g., Jones, Lignugaris/Kraft, & Peterson, 2007; Vaughn & Horner, 1997).

In the area of task difficulty, the theoretical ideas that have influenced this study most directly come from the work of psychologist Lev Vygotsky (1978), who forwarded important theories about task difficulty and learning. He posited that learning can actually lead development; tasks that require children to do things they cannot already do can actually spur their development in related areas. Vygotsky argued that learning takes place in a zone of proximal development (ZPD), which he defined as “the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers” (Vygotsky, 1978, p. 86). As

this definition suggests, Vygotsky was keenly interested in the role that social interactions play in individual learning. In the more than 70 years since his death in 1934, Vygotsky's ideas have continued to exert a strong influence on educational theory and practice, notably through the work of a number of scholars who have interpreted, tested, critiqued, and extended his theories. Two such scholars, Carol D. Lee and Peter Smagorinsky, have described some of the key assumptions about learners and learning that underlie Vygotsky's notion of the ZPD:

The capacity to learn is not finite and bounded. Rather, the potential for learning is an ever-shifting range of possibilities that are dependent on what the cultural novice already knows, the nature of the problem to be solved or the task to be learned, the activity structures in which learning takes place, and the quality of this person's interaction with others." (Lee & Smagorinsky, 2000, p. 2)

In other words, an individual's capacity for learning and the context in which learning takes place are inextricably linked.

Although Vygotsky was strongly interested in the role of language and literacy in individual learning and development, his work does not specifically address reading and text difficulty. However, his ideas have clear parallels to Betts' (1946) framework of reading levels, particularly the idea of the instructional level, which includes texts that an individual can read and understand with some form of assistance. In addition, the concept of a zone of proximal development has been adopted directly by publishers of educational materials that advocate ability-readability matching, as in the popular program Accelerated Reader, which describes the ZPD as "the range of books that will challenge a child without causing frustration or loss of motivation" (Renaissance

Learning, 2007, p. 4). And finally, with its depiction of connections among learners and activities and its firm grounding in larger, social contexts, Vygotsky's notion of the ZPD shares some clear similarities with the RAND (2002) model of reading comprehension described earlier. As a result, although Vygotsky was not specifically interested in children's performance reading texts of certain levels in certain contexts, his ideas are still helpful in informing our notions of text difficulty.

In relation to this study, Vygotsky's ideas have significant implications. If we accept his assertion that learning takes place when a child engages in an activity that requires her to achieve beyond what she can do independently, then the very notion of independent reading may be problematic in some ways, depending on the presumed and desired outcomes of independent reading time. The view that moderately challenging tasks lead to learning assumes that the learner will be receiving some form of support or scaffolding from a "more capable" individual, such as an adult or a skilled peer. As Pressley and his colleagues explain:

The risk in giving students moderately difficult tasks is that sometimes a student may be stumped. Rather than leave the student to flounder when confronted with a task that he or she cannot do, the teacher can 'scaffold' him or her (Wood, Bruner, & Ross, 1976), providing enough support so that the student can begin to make progress... Students can almost always have success with moderately difficult tasks with sufficient support – that is, with the kinds of tasks that allow them to see they can solve challenging problems and come to understand challenging material. (Pressley, Dolezal, Raphael, Mohan, Roehrig, & Bogner, 2003, p. 24)

If learning takes place in situations that include moderate challenge, social interactions and skilled assistance, then any potential benefits of reading more difficult texts may be called into question, particularly when the reading is done independently and without direct, instructional support. If a student is left to his own devices when it comes to choosing and reading a text, what forms of support are available to help him negotiate the situation and make sense of the text, especially a difficult text? Lee and Smagorinsky (2000) offer some insights into this issue, arguing that learning is “inherently social, even when others are not physically present” and that “language becomes the primary medium for learning, meaning construction, and cultural transmission and transformation” (p. 2). These two statements suggest the possibility that the language of the texts themselves may serve as a medium for learning and that an author may even be able to serve as a more capable other by providing scaffolds for understanding within the text. These ideas and others will be explored further in Chapter 5 of this dissertation.

Although it appears possible that textual scaffolds may be able to aid readers in making sense of difficult texts, the practice of classroom-based independent reading has faced both theoretical and empirical challenges. Some scholars have argued based on principle and on supporting evidence that academic feedback is essential for improving achievement (e.g., Gage & Berliner, 1992; Zahorik, 1987). In addition, although studies have found that recreational reading amount in general contributes to reading achievement (e.g., R. C. Anderson et al., 1988), experimental studies have thus far failed to find a positive effect on reading achievement for classroom-based independent reading as compared to traditional reading instruction (e.g., National Reading Panel, 2000). Despite these challenges, however, independent reading remains a common and popular

practice in elementary classrooms. One of the aims of this study is to find out more about students' experiences with difficult reading tasks in the absence of instructional scaffolding and social interaction.

### *Motivation to Read*

As with the previous construct of text difficulty, I lead into the discussion of motivation to read with a brief discussion of motivation more generally, focusing on one specific approach that is especially relevant to this study: the expectancy x value model. This model is based on the more formal theoretical work of a number of scholars in the area of achievement motivation, dealing with individual differences in the tendency to achieve or to pursue success (e.g., Atkinson, 1958; Atkinson & Raynor, 1974; Eccles, 1983; Spence & Helmreich, 1983). According to expectancy x value theory, individual behavior, specifically the tendency to achieve, is the product of expectations of success or failure and dispositional tendencies toward approaching success or avoiding failure, as determined by a combination of intrinsic and extrinsic motives and a mixture of affective and cognitive factors (Spence & Helmreich, 1983). In other words, a person's decision to undertake a particular task will be influenced by her general tendencies to desire success and to avoid failure, and by how probable she perceives success or failure at the task at hand.

In this way, the theory depicts behavior as the product of both relatively stable personality characteristics and situational perceptions. As such, people's choices and behaviors related to achievement are a complex mixture of "interacting motives that vary in strength and saliency across individuals, and within individuals, across situations" (Spence & Helmreich, 1983, p. 17). Additionally, Eccles (1983) asserts that in

considering expectations of success and the value of achievement, individuals act based not on the reality of their experiences, but on their perceptions of them. She explains:

...it is not reality itself (i.e., past successes or failures) that most directly determines children's expectancies, values, and behavior, but rather the interpretation of that reality. The influence of reality on achievement outcomes and future goals is assumed to be mediated by causal attributional patterns for success and failure, the input of socializers, perceptions of one's own needs, values, and sex-role identity, as well as perceptions of the characteristics of the task. Each of these factors plays a role in determining the expectancy and value associated with a particular task. Expectancy and value in turn, influence a whole range of achievement-related behaviors, e.g., choice of the activity, intensity of the effort expended, and actual performance (p. 79-81).

The last sentence in this quotation hints at an important application of the expectancy x value model, which is to help predict the difficulty level of tasks that an individual is likely to choose to pursue. Studies have generally found that people prefer tasks at an intermediate level of difficulty (e.g., Weiner, 1972).

For the purposes of this study, I have chosen to focus on a more general expectancy x value model, advocated by Brophy (2004), which is based on the more formal theory explicated above but is intended to apply to a broader range of learning situations. As Brophy (2004) explains, "applications of the more specific theory are usually limited to achievement situations that call for meeting clear standards of excellence" (p. 22). Because this study focuses on classroom-based independent reading, which is not a typical achievement situation, the broader expectancy x value model is an

appropriate fit. It allows us to interpret individuals' behavior in a situation for which there are no set expectations of success and no clear definitions of failure. Expectations of success are therefore tied to more fluid and personalized notions of success. The more general model also allows for a broader conception of the term "value," which can include not only the degree to which students value successful outcomes, but also the degree to which they value the process of engaging in the task (Brophy, 2004).

In particular, the expectancy x value model as an overarching framework can make some specific, important contributions to investigations of students' reading behaviors. First, it reminds us of the importance of focusing not just on students' actual reading performance but also on their perceptions of their own reading abilities. Some research suggests that this distinction is especially important when dealing with young readers, who tend to be overly optimistic about their abilities (Chapman & Tunmer, 1995; see also Pressley, 2006, for a review). In the context of this particular study, the expectancy x value model shaped the methodological decision to include numerous opportunities for students to express their own ideas about their reading abilities and performance, because this information is viewed as centrally important to understanding reading behaviors such as text choice, task persistence, and task enjoyment. As will be seen in later chapters, the theory also informs the analysis and interpretation of data related to motivation and text difficulty.

Second, the expectancy x value model suggests two different roles that motivation might play in students' reading: 1) as an input, in terms of ways that it influences students' decisions about what texts to select and read; and 2) as an outcome, in terms of ways that it is affected by students' reading experiences with different texts. In the first



instance, motivation as an input could lead students to choose texts that offer a reasonable chance of success and that carry some value for the reader, such as interesting information or desired social interactions. The difficulty of a particular text would likely influence students' expectations of success, and it may also affect the value they attribute to the text. For example, for the first grade students in Bass' (2006) study, the challenging nature of the chapter books they read led students to value them as status symbols. In the second case, motivation as an outcome may be seen in the ways that a reading experience shapes students' sense of competence to interact with future texts and their evolving beliefs about the value of reading more generally. The role of text difficulty in these situations may be to mediate the connection between reading and motivation, although the nature of this relationship is unclear. For example, Betts (1946) described difficult texts as leading to students feeling "worried and unhappy" (p. 448), while Pehrsson (1994) has suggested that challenging texts are actually more effective than easier texts at helping children develop a sense of accomplishment and a desire to engage in additional reading experiences. The key to this relationship between reading experiences and motivation as an outcome may be the degree to which the reader views the reading experience to have been successful, given that success motivates by increasing students' self-efficacy (see Pressley et al., 2003, for a review).

This last point suggests the importance of engaging students in reading experiences that will be successful and therefore motivating. However, it is important to note that the very idea of what makes a reading experience successful is still up for debate. While teachers and other adults may define successful reading as accurate and fluent, with good comprehension, young readers may have different ideas. In particular,

they may or may not have the same perceptions of a text's difficulty – and their own “success” at reading it – as a teacher might determine through traditional methods and assessments. This study aimed to examine motivation in both of these roles – as an input and as an outcome – in order to learn more about why students chose the texts they did and what role text difficulty played in determining whether a student judged a reading experience to be successful and motivating. To this end, the expectancy x value model of motivation influenced the design of the study by suggesting the inclusion of measures of students' perceptions of reading ability, text difficulty, and the value of reading.

### *Summary*

The preceding sections have offered some broad theories in the areas of reading comprehension, text difficulty, and motivation to read. Each of these ideas has informed this study, in terms of the research questions, study design, and data analysis. The RAND model of reading comprehension draws our attention simultaneously to the reader, the text, the activity, and the sociocultural context. Vygotsky's notion of a zone of proximal development requires us to look both at text difficulty and at the textual, interactional, and instructional scaffolds that may support students in their attempts to read difficult texts independently. The expectancy x value model of motivation requires us to look at the ways that motivation influences and is influenced by students' text choices and reading experiences. Individually, these different theories have specific implications for the design and implementation of the study; taken together, they help form a foundation that grounds the work in a complex network of relationships between students, texts, and the varied activities of the classroom setting.

## CHAPTER 3: METHODS

### Overview

To assess student reading ability and motivation to read, all students in the sample completed a standardized reading test and a written survey of reading motivation. Written reading logs were used to collect information about students' daily reading choices and reading experiences. Reading test scores and information from the logs were then used to help identify occasions in which students appeared to have chosen texts that may have been at their frustration level, based on an apparent mismatch between text readability and student reading ability. For each of these identified occasions, students then participated in a series of short, one-on-one assessments and a follow-up interview. The text-specific assessments provided information about students' oral reading accuracy and comprehension for the difficult texts they had read. The interviews gathered additional qualitative information about the nature of students' experiences with their chosen texts, including their understanding of them and their enjoyment of them. In order to have a point of comparison, I also identified a parallel occasion for each student in which he or she appeared to have selected an independent-level text. Each of the text-specific measures was then used for the easier text as well. In some instances, when an independent-level match was not identified from the set of texts a student had chosen, the student was asked to choose from a small set of easier texts provided by the researchers.

To address the five research questions outlined earlier, this study used a mixed methods design and multiple data sources. Johnson and Onwuegbuzie (2004) have defined mixed methods research as "the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches,

concepts, or language into a single study” (p. 17). This approach was selected because it seemed to be an appropriate fit for the research questions under consideration, which call for a combination of deductive and inductive work. The quantitative data sources offer opportunities for statistical analyses and hypothesis testing, while the qualitative data help to flesh out the numbers with more detailed information and the potential for describing some illustrative cases. The remainder of this chapter offers more detailed descriptions of the sample, data sources, and procedures.

### Study Context and Sample Description

The focus of this study was on second grade students’ text choices during classroom-based independent reading time. The target grade level, the study context, and the sample size were chosen based on several determinations related to the content of the study.

#### *Grade Level*

Second grade classrooms were chosen for several reasons. Because oral reading accuracy is an important part of this study, it was important that students be old enough to be good decoders, but still not perfect decoders. Second grade was therefore chosen because it is an important transitional year in students’ reading skill development, with many children moving from word-by-word decoding toward more fluent reading (e.g., International Reading Association, 1998; Juel, Griffith, & Gough, 1986; Pressley, 2006). Because choice is also a central focus of this study, it was important to work in classrooms with a diverse range of available texts. Second grade fit this criterion as well, because it often marks a transitional point in text complexity, as readers move from

simple decodable or predictable books to complex picture books, detailed informational texts, and even chapter books.

### *Classroom Context*

The decision to situate the study within the context of daily classroom practice was driven by a desire to ensure that the collected data would reflect actual classroom activities and experiences as much as possible. In addition, because one purpose of the study was to examine students' motivations for choosing and reading particular texts, it was important to ensure as high a degree of authenticity of choice as possible. A number of specific steps were taken to limit constraints on students' text choices and to increase the likelihood that text choices would reflect students' actual text preferences. First, classrooms were selected through purposive sampling, such that each participating classroom afforded its students frequent opportunities of "free" choices of materials for independent reading. To meet this goal, a set of classroom selection criteria was created to address three factors: 1) classroom library inventory, 2) reading time, and 3) choice environment.

Regarding library inventory, in order to be selected, classrooms had to have a minimum of 25 titles per student and a variety of different types of texts according to genre, format, and difficulty. This stipulation was intended to guarantee that students would be able to find texts that were genuinely appealing to them. For reading time, classrooms were only selected if they offered students between 20 and 30 minutes of self-selected, independent reading time at least three times each week. This requirement was put in place to ensure that students participating in the study, which took place in the spring, would be familiar and comfortable with classroom independent reading routines

and procedures by the time data was collected. In terms of choice environment, teachers were asked to allow students to choose freely from the entire selection of available texts in the classroom library. By removing any existing guidelines or restrictions based on things like genre and difficulty, I hoped to reduce choice constraints to the limitations of the library inventory. It is important to emphasize here that the free choice environment itself was not the focus of the study. Rather, these steps were taken in the hopes of establishing an idealized context in which students chose from a diversity of texts, so that their choices closely approximated what they might have chosen in a similar or an even less restrictive setting.

### *Sample*

For several reasons, it was important that this study involve multiple classrooms. First, because social factors often play a role in individual students' text choices (e.g., Bass, 2006; Halladay, 2006), multiple classroom contexts were necessary for observing a range of text choices and behaviors. For example, because peers can influence the choices of their classmates, a single classroom may not reflect much diversity in reading behaviors. Second, because characteristics of classroom libraries – including the texts they contain and the way they are organized – also impact individual choices, it was necessary to have several different libraries represented. Third, because gender has also been shown to be an important factor in student reading behavior (e.g., Childress, 1985), the sample needed to be large enough to allow for gender to be considered as an independent variable. With the sample composed of occasions in which students chose difficult texts, the goal was to have a sample of 40 occasions, each for an individual student, which would offer approximately 20 occasions per gender. A classroom sample

size of five classrooms was estimated to be sufficient for identifying the target number of 40 occasions.

To begin recruiting the five participating classrooms, several potential school districts were identified, largely based on their geographic proximity to the University. Information about the study was sent via e-mail to teachers, principals, and district administrators with whom I had had previous contact through research projects or field supervision experiences. This approach was appropriate because the sample of classrooms was meant to be purposive, with classrooms required to meet the set of criteria described earlier in this chapter. These selection criteria were therefore the only considerations used to identify classrooms.

When a teacher expressed interest in participating in the study, I made a preliminary visit to the classroom to meet the teacher and to get more information about the classroom library and about the structure of the students' classroom independent reading time. These preliminary visits served the dual purposes of giving the teachers more information about the study and checking to see whether the classroom met the selection criteria. At each visit, I completed an informal inventory of the classroom library, taking detailed notes on the number, diversity, and organization of available texts. I made preliminary visits to five classrooms, all of which met the criteria and qualified for participation the study. This high rate of recruiting success is likely due to the fact that recruitment materials had included the list of selection criteria. Teachers may have only responded if they believed they would qualify for participation in the study.

The five participating second-grade classrooms were located in four different elementary schools, representing three school districts on the outskirts of a mid-size

Midwestern city. Although this study is not directly concerned with demographic variables other than gender, some contextual information seems appropriate. Table 3.1 offers some school-level data for student socioeconomic status and English Language Arts achievement testing scores.

Table 3.1. Race, socioeconomic status, and achievement data by school.

Classrooms	Enrollment <sup>a</sup>				Free/Reduced Lunch <sup>b</sup>	ELA Proficiency <sup>c</sup>
	White	Black	Hispanic	Asian/Pacific Islander		
A & B	92.4%	1.6%	3.8%	2.2%	29.3%	80.8%
C	91.7%	1.8%	4.0%	2.5%	24.4%	80.0%
D	30.5%	18.4%	8.5%	42.6%	38.1%	75.9%
E	65.6%	5.4%	3.9%	25.1%	12.4%	95.5%

<sup>a</sup> Enrollment ethnicity data (National Center for Education Statistics, 2008) are from the Common Core of Data (CCD) for the 2005-2006 school year. <sup>b</sup> Free and Reduced lunch data (Michigan Department of Education, 2008a) reflect the percentage of enrolled students in the 2007-2008 school year who qualified for free or reduced lunch. <sup>c</sup> ELA Proficiency (Michigan Department of Education, 2008b) represents the percentage of 3<sup>rd</sup>-5<sup>th</sup> grade students in the school who achieved proficiency on the English Language Arts section of the Michigan Education Assessment Program (MEAP) for the 2007-2008 school year.

Parents and guardians of all students in the five participating classrooms were asked to provide informed consent for their children's participation in the study. Out of the 103 students who received consent forms, 72 returned signed consent forms and were included in the initial sample, which consisted of 36 boys and 36 girls. Two students, both girls from the same classroom, were later removed from the sample because they were English Language Learners whose limited proficiency in English made the one-on-one assessments nearly impossible. The final sample, then, consisted of 70 second-grade



students, of whom 36 were boys (51.4%) and 34 were girls (48.6%). Parents of several of the students declined to provide their child's age, but for the students for whom age data was available ( $n = 63$ ), ages ranged from 87 months to 107 months ( $M = 95.51$ ,  $SD = 4.45$ ). No other demographic information was collected for individual students. Because of variations in class size and return rate, the final sample of 70 students was not distributed evenly across the five participating classrooms. Table 3.2 shows the sample distribution across classrooms.

Table 3.2. Sample distribution across participating classrooms.

Classroom	Consented Students			Total Students	% participating
	Boys	Girls	Total		
A	11	8	19	24	79.2%
B	9	12	21	24	87.5%
C	5	4	9	18	50.0%
D	4	3	7	17	41.2%
E	7	7	14	20	70.0%
Total	36	34	70	103	68.0%

### *Classroom Libraries*

Overall, the five classroom libraries were quite similar in their contents and organization, diminishing the likelihood that these aspects of the classroom setting would lead to any significant classroom-level effects on students' reading choices. The selection criteria were designed to achieve a level of uniformity across classrooms, thus limiting the number of factors that could be seen as contributing to differences in student behavior. Although the classroom context is certainly an important part of student

behavior and student learning, data analysis for this study did not focus on differences between classrooms. Because of the high degree of similarity among libraries in the participating classrooms, I describe them as a group, focusing on their clear similarities rather on their subtle differences.

*Inventory.* All five classroom libraries met the minimum criterion of 25 titles per student. Exact numbers of texts were not counted, but detailed estimates suggest that library inventories ranged from a low of approximately 720 texts to a high of more than 1430 texts. One teacher had recently received a grant of \$1,000 from her school district's education foundation to purchase new books, labels, and bins for her classroom. In another class, the students did their daily reading time outside of the regular classroom, in a large multipurpose room with a vast collection of texts shared by all of the second and grade classrooms in the school. The largest library was in the classroom of a veteran teacher who admitted that her sizable collection was the product of decades of gradual work and steady accumulation.

In addition to offering a large number of texts, all of the classroom libraries also offered a broad range of reading materials, including different formats, genres, and levels of difficulty. Students could choose from a wide selection of picture books, transitional readers, chapter books, and informational texts. Libraries also contained smaller selections of materials such as simple decodable texts, poetry collections, rhyming books, magazines, and reference books. Two classrooms had some graphic novels and comic books, and one even had a small collection of newspapers and greeting cards.

In all of the classrooms, the sizable collections of diverse texts were grouped largely by genre, author, topic, theme, and series. For example, books were generally

arranged in bins with category labels such as “Mystery,” “Dr. Seuss,” “Science: Biomes and Plants,” “Celebrations,” or “Magic Tree House.” Most of the classrooms also had at least one area where thematically linked texts were displayed on a rotating basis, generally in connection to topics of classroom study, such as a content area topic (e.g., frogs, soil) or an author or genre study (e.g., fairy tales, the Amelia Bedelia series). Several of the classrooms also had some portion of the texts organized either by readability level or by a reasonable proxy of readability. For instance, one classroom had two bins of “Level 2” books, which were leveled readers written at a second grade reading level; another class had a small set of shelves with bins labeled by the color-coded levels of the Reading Recovery leveling system (Clay, 1993b); and a third classroom had two bins labeled “quick reads,” which were filled with shorter texts that could be read in one sitting. On the whole, the classroom libraries were highly similar in terms of the number, diversity, and organization of available texts.

*Reading Time.* All of the classrooms offered regular periods of independent reading time, usually as part of a daily reading workshop block. Because it was part of a larger instructional block, the actual reading time tended to vary a bit, depending on how much time was spent on other block components, such as minilessons and guided reading groups. One classroom offered 25 minutes of silent reading time directly after lunch. On average, though, based on informal observations and conversations with teachers, students in all of the classrooms spent approximately 20 minutes each day on self-selected, independent reading.

*Choice Environment.* All five teachers were asked to remind students at the beginning of each independent reading period to choose freely from the collection of

available texts, in order to increase the chances that students' choices reflected their actual preferences as much as possible. All five teachers agreed to offer their students unlimited free choice, within the obvious constraints of the classroom library inventory. For some, this approach was very consistent with their daily routines. For example, one teacher said that she always offered free choice and that she downplayed the role of text difficulty in favor of factors like interest and prior knowledge, encouraging students to actively monitor their understanding as a way to evaluate a text's appropriateness for them. Another teacher also offered free choice, commenting that she did not teach her students book choice skills or offer them suggestions for evaluating a text's appropriateness. A third teacher used a reading workshop approach and offered her students free choice from the classroom library in addition to direct instruction in book selection, encouraging students to choose based on a range of factors including interest, prior knowledge, and difficulty. For other teachers, the practice of allowing free choice represented a bit of a departure from their regular routines. One teacher explained that her students "mainly pick by high interest" and that they use the 5-finger method<sup>1</sup> "usually but not always." Another teacher also used this method – her classroom had a posted list of guidelines for choosing a "just right" text, which included an instruction to "choose books that you think you will enjoy. Put books back that are too hard (5 finger test)." In addition, her classroom contained another sign near the bookshelves that offered sets of questions to help students decide if books were "too easy," "just right," or "too hard."

The list of questions for "too hard" read as follows:

---

<sup>1</sup> This 5-finger method is a practice commonly used in elementary classrooms to help students determine for themselves whether or not a book is too difficult for them. The method involves reading a sample page from a text while keeping track of any unknown words you encounter by counting them on your fingers. If there are more than 5 unknown words on a single page, then the book is judged to be too difficult.

Are there more than 5 words on a page that you don't know?

Are you confused about what is happening in this book?

When you read, does it sound pretty choppy?

This teacher also mentioned that when students picked books that were “too challenging” based on the 5-finger rule, she helped them find a different book to read. On the whole, teachers’ practices were highly similar in the degree of free choice they typically offered their students, and all five teachers agreed to encourage free choice during the data collection period. There were some slight differences in the amount of guidance teachers offered to their students, particularly in the level of emphasis they placed on text difficulty. Although all five teachers had similar practices and agreed to offer and support free choice during the course of this study, it is still possible that some students were influenced by the standard procedures that they had learned and practiced during the earlier part of the school year.

### Data Collection

#### *Reading Ability*

To get a general measure of reading ability, each student completed two subtests from the Level 2 Gates-MacGinitie Reading Test (GMRT-4; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2002), a norm-referenced assessment that can be administered in a group setting. The Level 2 GMRT-4 contains three subtests: Word Decoding, Word Knowledge, and Comprehension. For this study, only the Word Decoding and Comprehension subtests were used, because those are the constructs that align most closely with standard reading level criteria, as seen in Betts’ original framework and in commercially-available IRIs. The Word Decoding subtest consists of 43 items that

require students to look at a picture and then choose one of four orthographically similar words that best matches the picture. Students were given 20 minutes to complete this subtest. The Comprehension subtest consists of 39 items that require students to read short stories and non-fiction passages and then choose one of three pictures that best matches the corresponding written segment. Students were given 35 minutes to complete this subtest. The GMRT-4 has been shown to have good reliability and validity, including internal consistency reliability of .97 for the Level 2 tests (Johnson, 2005). Content validity has been established through item response methods, and concurrent validity is inferred from high score correlations with the Third Edition of the GMRT (MacGinitie & MacGinitie, 1998), which has been shown to be highly correlated with a number of other reading tests (Johnson, 2005; MacGinitie & MacGinitie, 1989). Raw scores can be converted to Grade Equivalent scores or to Lexile<sup>2</sup> readability scores, facilitating comparisons between student reading ability and text readability.

When signed consent forms started being returned, arrangements were made for administering the GMRT-4. The Level 2 version of the GMRT-4 is available in two different forms (S and T), and it has alternate form reliability of .95 (Johnson, 2005). Because this reliability is relatively high, it was not necessary to balance the test form randomly across the sample, so all students within classrooms took the same form of the test, either S or T. Teachers were given the choice of either administering the GMRT-4 subtests themselves or having me visit their classroom to administer it. Three of the teachers chose to have me administer the test; in the other two classrooms, the classroom

---

<sup>2</sup> The Lexile framework is a single scale that can be used as a quantitative measure of both text readability and individual reading ability. The Lexile measure of text readability is based on two features of a given text: semantic difficulty, as measured by word frequency (how rare or how common the words in the text are); and syntactic complexity, as measured by sentence length (Lennon & Burdick, 2004).

teachers administered the test themselves and were instructed to follow the standardized administration procedures. In one of these cases, the teacher chose to administer it because it was easier for her to fit it into her busy schedule. In the other instance, the teacher administered the test as part of her district's annual assessment plan. Because the students were going to be taking the test anyway, arrangements were made for that data to be released for the research project as part of the standard parental consent process. In the other four classrooms, I collected the completed test booklets from consented students and hand scored them using the booklet provided by the test publisher. Test data were recorded in a database as raw scores, grade equivalent scores, and Lexile levels.

### *Reading Motivation*

All students completed the survey portion of the Motivation to Read Profile (MRP; Gambrell, Palmer, Codling, & Mazzoni, 1996). The survey contains 20 items designed to assess students' self-concept as readers and the value they place on reading. Each of the 20 items consists of a short prompt and four choices, arranged on an ordinal scale. For example, one of the self-concept items reads "I am..." with the choices "a very good reader," "a good reader," "an OK reader," and "a poor reader." No information on internal reliability is available for the MRP survey, but its validity has been determined through comparisons of survey responses with student interview data and through statistical tests of relationships between survey responses and student achievement and between survey responses and grade level (Gambrell et al., 1996). Both raw and percentage scores can be calculated for the individual subscales and for the survey as a whole.

As with the GMRT-4, teachers were given the choice of either administering the MRP survey themselves or having me visit their classroom to administer it. All but one of the teachers chose to administer the survey on their own, so they were given copies of the surveys and a copy of the administration directions that were published with the original instrument and that they were asked to follow. As recommended by the profile's creators, all of the survey items were read aloud to students in a group setting, and students indicated their responses by checking the appropriate boxes on their paper copy of the survey. After surveys were completed, I collected them from consented students and scored them using the guidelines provided with the published instrument (Gambrell et al., 1996). Survey data were recorded in a database as individual item responses and as subscale and total scores.

### *Reading Logs*

All teachers were given copies of a daily log in which their students were to keep track of the texts they chose and the texts they read during classroom-based independent reading time (see Appendix A). This log was comprised of two sections. The first focused on texts chosen, and the second focused on texts actually read.

1. *Chosen texts.* Students completed this portion of the log each time they chose a text that they planned to read during independent reading time. Students wrote down the title of the text and explained their choice by providing a written response to the prompt, "I chose this book because...". This question was intended as one way to get at the important issue of why students choose texts for independent reading.



2. *Read texts.* Students completed this second portion of the log after they finished reading a text. For each text, they indicated their perception of its difficulty for them by checking a box next to one of five options: too easy, kind of easy, just right, kind of hard, or too hard. The log also included a scale item that prompted students to indicate their enjoyment of the text by filling in from one to five stars and by adding comments to explain their rating.

Teachers began using the reading logs with their students in April. Teachers were responsible for providing initial instruction in how to fill out the logs. They were also asked to remind their students to fill out the provided logs as they chose and read texts during their independent reading time. At the end of the data collection period, which lasted approximately nine weeks, classroom teachers collected their consented students' reading logs and submitted them. Individual reading log entries were typed into a database for further analysis.

#### *Identifying Occasions*

Once the GMRT-4 and the MRP were completed, another graduate student and I began visiting the classrooms to identify potential occasions in which students had chosen to read frustration-level texts. Having already scored the GMRT-4 test booklets, we were able to enter the classrooms with a measure of reading ability in hand for each participating child. On each classroom visit, we approached consented students individually and asked them to show us their reading log pages. We reviewed the log entries, looking for occasions in which they may have chosen a frustration-level text. To identify these occasions, we considered several pieces of information. As much as possible, we used their reading ability scores (translated into Lexile levels and grade

equivalent scores), the text titles in the reading logs, and our own existing knowledge of the texts they had chosen. For example, one student had scored at a 1.3 grade level on the word decoding subtest and a 1.8 grade level on the comprehension subtest, which equates to a Lexile level of 30L. Based on this information and on my knowledge of Magic School Bus books, I guessed that his choice of *The Magic School Bus at the Waterworks* (Cole, 1988) – which has a Lexile level of 660L and an ATOS grade equivalent of 3.7 – might be a frustration-level text for him. In some instances, when we were unfamiliar with the texts listed in a student's log, we asked him or her to bring a few texts to look at so that we could flip through the pages get general estimates of readability and potential difficulty for the student. Once a student and a text had been chosen, the researcher took the student to a quiet area and began the series of four one-on-one tasks: oral retelling, running record, comprehension questions, and interview. The administration procedures for each of these tasks are described in some detail below. Additional information about scoring and analysis of the individual measures is included in the upcoming Data Analysis section of this chapter.

### *Oral Retellings*

To assess students' comprehension of individual texts, students were asked to give oral retellings of some of the texts they had recently chosen to read during classroom-based independent reading time. As mentioned earlier, the oral retellings were conducted both for texts that appeared likely to be frustration-level texts and for matched occasions with texts that appeared likely to be easier than frustration level.

Oral retellings were completed prior to the running records and comprehension questions, which were based on shorter passages selected from the larger texts, so that

reading the passages would not influence students' retelling of the text as a whole. The administration protocol for the oral retellings was modeled after retelling guidelines in published IRIs. It was designed to be an open retelling procedure with only a few general prompts for additional information. The researcher began by saying, "First, I am going to have you tell me what you remember from reading this book. I'm going to use a tape recorder to help me remember what you say. Is that okay?" After the student indicated that he or she understood, the researcher said, "Tell me everything you can remember from this book." The researcher then listened as the student responded. When the student first paused or indicated that he or she had finished, researchers were directed to offer the prompt, "What else do you remember?" Again, the researcher listened, and the next time the student paused or indicated that he or she had finished, the researcher offered the final prompt, "Is there anything else you would like to add?" After the student made any final additions to his or her retelling, the researcher moved on to the next procedure, which was the running record of oral reading fluency. Oral retellings were audiotaped for later transcription, scoring, and analysis.

### *Running Records*

Running records were used to assess students' oral reading accuracy and to determine whether their selected texts were at their independent, instructional, or frustration level. Running records are a common method of assessing oral reading accuracy, in which an individual reads aloud from a text while the assessor compares the oral reading performance to the original text and records any deviations, or miscues. Because students were being assessed on texts they had already read, the running records tested students' ability to read previously read passages rather than unfamiliar passages.

According to Betts (1957), this method – although not ideal – is acceptable, since “a fairly satisfactory inventory of reading performance can be made with materials which the child has ‘read’ before” (p. 455). In fact, Betts’ original reading level criteria (1946) were based on students reading silently before they read orally (Jongsma & Jongsma, 1981). Additionally, this method is consistent with some assessment procedures, which either require (Flynt & Cooter, 2001) or allow (Bader, 1980; Burns & Roe, 1999) students to read test passages silently to themselves before reading them aloud. The running records were audiotaped for later analysis, both for actual scoring and so that inter-rater reliability could be established. The scoring system for this measure is described in more detail in the upcoming Data Analysis section of this chapter.

After a student finished his or her oral retelling, the researcher looked through the target text and selected a passage to be used for the running record and for the passage-specific comprehension questions. Passages were selected according to the following set of guidelines: 1) passages should be chosen somewhat randomly from the approximate middle of the target texts; 2) passages should not begin mid-sentence or mid-paragraph, but they may begin mid-page or mid-chapter; and 3) passages should be approximately 150 words in length. In the few cases in which the entire text was shorter than 150 words, the whole text was used. These guidelines were intended to ensure that selected passages were representative of the texts as a whole, contained whole units of meaning, and were long enough to provide sufficient information about oral reading behaviors. They were also designed to be consistent with commonly used procedures. For example, Fountas and Pinnell (1996) recommend using passages of approximately 150 words; Clay (1993a) advocates between 100 and 200 words. Additionally, a sampling of 24 second grade level

passages from 6 different commercially published IRIs (Applegate, Quinn, & Applegate, 2008; Burns & Roe, 1999; Flynt & Cooter, 2001; Johns, 1997; Stieglitz, 2002; Woods & Moe, 2007) revealed an average passage length of 145.25. This sampling suggests that the passage length used in this study is roughly equivalent with common recommendations and assessment procedures.

After selecting a passage, the researcher read the following directions aloud to the student:

Now I would like you to read part of this book to me out loud. I will show you where to start and stop reading. As you read, try to pronounce the words as best you can. Also, try to remember what you are reading. When you have finished, I will ask you some questions about what you read.

These directions were intended to encourage students to focus both on the task of pronouncing individual words and on the task of deriving meaning from the text. Given the research that teacher directions can influence student comprehension and reading performance (Pehrsson, 1974; Furniss & Graves, 1980; Jongsma & Jongsma, 1981), these explicit instructions were believed to be consequential. Especially because the oral reading procedure preceded the comprehension measure, it seemed important to direct students specifically to understand and remember, so that they would not focus too heavily on the oral reading as a performance. At the same time, it is important to acknowledge that even these directions may have had some effect on students' reading during the one-on-one assessments.

Next, the researcher pointed to the beginning of the selected passage and asked the student to begin reading. As the student read, the researcher listened, counted words

on a tally sheet, and audiotaped the reading. In accordance with common IRI procedures, researchers offered no prompts or assistance other than supplying words if the student asked for help or paused for more than about 5 seconds (Applegate et al., 2008; Burns & Roe, 1999; Flynt & Cooter, 2001; Stieglitz, 2002). Asking for help was considered to include either an actual verbal request for help or a non-verbal appeal, such as stopping and looking at the researcher with a questioning look. Researchers used the tally sheets to count words up to 150; after reaching 150 words, reading continued until the student also reached the end of the sentence, and then the researcher indicated that the student could stop reading.

### *Comprehension Questions*

Sets of 6 comprehension questions were generated to assess students' understanding of the passages they had just read as part of the running record assessment. Questions were designed to be passage-specific, and they were generated based on a framework derived from an informal review of several published IRIs and in accordance with Valmont's (1972) guidelines (see Appendix B for a copy of the framework and guidelines). This process of generating comprehension questions was designed with two goals in mind: 1) to create questions that were closely matched to the type of questions teachers frequently use in assessing their students' reading comprehension; and 2) to create questions that would accurately assess the reader's comprehension of the target passage. Students' responses to these questions were used as a complement to the oral retellings, to provide convergent evidence of students' comprehension of their chosen texts.

When a student finished reading the running record passage aloud, the researcher closed the book and gave the following instructions:

Now I am going to ask you a few questions about the section you just read aloud to me. For each question, just give me your best answer. If you don't understand a question or if you need me to repeat it, just ask. Okay?

After the student indicated that he or she understood the directions, the researcher then quickly reread the running record passage and generated a series of approximately 7 passage-specific comprehension questions. Some texts were either too short or too simple (or both) to support this number of questions; in those cases, a smaller number of questions – ranging from 4 to 6, depending on the text's content – were asked. The final goal was to have 6 good, scorable questions for each passage, so an extra question or two in this data collection stage allowed for greater selectivity during the scoring stage. The question selection process is described in more detail in the upcoming Data Analysis section.

As mentioned above, the process of generating questions was guided by two different tools: 1) a set of question shells derived from a review of commercial IRIs; and 2) a list of criteria for good comprehension questions (adapted from Valmont, 1972). Both of these tools are included in Appendix B. The question shells were developed based on an analysis of the types of questions used in a variety of commercial IRIs. The purpose in developing and using these question shells was to promote consistency across and texts and to ensure that generated questions would be highly similar to those used in published IRIs. The list of criteria for good questions included rules such as, “avoid questions that are answerable from pictures in the text” and “ask passage-dependent

questions that cannot be answered from prior or general knowledge.” The method of generating questions on the fly is difficult to do with any degree of consistency, especially given the multiple considerations that guide question formation. To improve the quality and reliability of the comprehension questions, researchers practiced the question generation process repeatedly on a variety of practice texts until the question shells and criteria were internalized and the process became more automatic (for a sample set of comprehension questions, see Appendix C). When working with students, researchers read each question aloud and then listened to student responses. If necessary, researchers probed for clarification or additional information by saying, “Can you tell me more about that?” This probe was only used once for each question. All responses were audiotaped for later transcription and analysis. The researcher also kept the chosen text so that the selected passage used for the running record and the comprehension questions could be typed into a laptop computer. It was necessary to have a written record of text passages so that students’ responses to the passage-specific assessments could be compared to the words and meaning of the actual text. In addition, information about each text was recorded so that Lexile scores could be obtained through the Lexile website ([www.lexile.com](http://www.lexile.com)), which offers Lexile scores through both an extensive database of texts and an online tool for analyzing text samples.

### *Interviews*

One-on-one interviews with students were used to gather additional information about students’ experiences with their chosen texts (see Appendix D). Questions focused on reasons for choosing the target text, previous experiences with the target text (e.g., number of times read), perceptions of its difficulty, perceptions of reading performance,



and enjoyment of the target text. Student responses were audiotaped for later analysis. Information gathered through these interviews was intended to complement and extend students' reading log entries and to speak to the issues of reasons for choosing, perceptions of text difficulty, and enjoyment.

Upon finishing with the set of passage-specific comprehension questions, the researcher transitioned to the interview by saying, "Now I am going to ask you a few more questions about the book in general." This statement was intended to focus the student's attention back on the text as a whole rather than only on the short passage he or she had just read. The researcher then read the interview questions aloud and listened to student responses. All interviews were audiotaped for later transcription and analysis. At the end of the interview, the student returned to regular classroom activities.

#### *Identifying Matched Occasions*

After each day of data collection, I scored students' running records to determine whether each target text had been at an independent, instructional, or frustration level for the student who had read it. On subsequent visits, I continued reading with students until I had read with each student once, and then I began working with students for a second time, trying to find either independent-level matches for students who had previously chosen frustration-level texts, or vice versa (this matching process is described in more detail in Chapter 4). This method proved to be largely adequate for identifying the necessary frustration-level occasions. However, for a number of students, it was difficult to find a matched occasion in which they read an independent-level text, even after multiple trials. In these situations, researchers provided students with a small selection of easier books from which to choose. A list of provided texts can be found in Appendix E.

It is important to note that not all of these texts were offered as choices to individual students; each student was offered a choice of about 6 texts, which were estimated to be at that student's independent reading level. Students were given time to read these easier, provided texts before completing the one-on-one assessments.

The search for matching occasions resulted in a final set of 35 matched pairs, out of which 19 of the independent-level readings were from provided texts rather than from texts freely chosen during independent reading time. This method of providing books compromised the naturalistic intentions of the study, but it was essential to have a point of comparison for each difficult text. Since the focus of this study was on frustration-level texts, it seemed more important to the study that the difficult texts be authentic choices than that the independent level texts be authentic choices as well. Additionally, this method of providing easier texts was used only after researchers had already read at least twice with a given student.

## Data Analysis

### *Scoring*

*Oral Retellings.* Audio files of oral retellings were transcribed and scored against a general, 10-point rubric (see Appendix F), which was adapted from a similar tool used in Johns' (1997) *Basic Reading Inventory*. Rather than focusing on text elements particular to a certain genre, this general rubric included characteristics like sequence of ideas and overall accuracy and coherence. One might wonder why I chose this general approach in lieu of a genre-specific scoring method, which is used in a number of commercially available IRIs (e.g., Applegate et al., 2008; Goodman, Watson, & Burke, 2005; Woods & Moe, 2007). The reason was that the wide range of text types selected by

students made more specific rubrics virtually unworkable. For example, standard oral retelling rubric items for narrative texts include plot episodes and descriptions of main characters, but these items do not apply well to texts like Dr. Seuss's *The Foot Book* (Dr. Seuss, 1968), which is primarily a rhyming book with only a very loose narrative thread and no identifiable central characters. In addition, neither the narrative nor the expository rubric alone could be used to assess the retelling of hybrid texts such as those in author Joanna Cole's popular *Magic School Bus* series, which contain elements of both genres. And as a final example, neither genre-specific rubric was adequate for assessing students' retellings of poetry collections, which fall into neither category. Applying a genre-specific rubric that was an ill fit for some texts may have affected scores in such a way that they would correspond to text's characteristics rather than to student comprehension. In other words, a genre-based rubric may have measured the degree to which the text fit the rubric's profile of genre elements better than it would have measured a student's actual understanding of the text. For these reasons, I chose to use a more general scale that could largely be applied to a text of any type.

*Running Records.* For each of the assessed texts, a score sheet was created that included the typed passage and spaces set aside for tallying different types of oral reading errors. The score sheets also included a 4-point scale for evaluating prosody, a 4-point scale for estimating the degree to which a student's oral reading errors affected the meaning of the passage they read aloud, and a section for calculating a student's reading rate based on number of words read and time spent reading, although these measures were not analyzed for this study. For a sample running record score sheet, see Appendix G.

Establishment of scoring rules for running records was somewhat difficult to achieve, given the fact that there is currently no consensus around exactly how oral reading performances should be evaluated (see Appendix H for a matrix comparing a variety of scoring guidelines for running records and IRIs). In particular, there is considerable difference of opinion regarding what should count as an error and whether all errors should carry equal weight. Given this lack of consensus, the scoring rules for the running records in this study were created with two sometimes competing goals in mind: 1) assessment validity and 2) consistency with commonly used methods. For every scoring decision, I believed that it was important to be able to justify it as contributing to an accurate measurement of a student's reading performance. At the same time, however, because this study is in some ways a test of the utility of traditional procedures and criteria, I felt that it was important not to stray too far from the actual methods used by popular IRIs. In several situations, these competing goals led to compromises. For example, some scholars contend that self-corrections and repetitions are not errors at all, but are instead indicators of reading skill – signs that the reader is actively monitoring comprehension and making multiple attempts to make sense of the text (e.g., Applegate et al., 2008; Clay, 1985; Flynt & Cooter, 2001). Based on this information, it was tempting to choose not to score self-corrections or repetitions as errors. However, most commercially published IRIs include either repetitions or self-corrections, or sometimes both, in their lists of scorable errors. To remain consistent with common IRI methods, I therefore decided to count one but not the other, and I chose to count self-corrections as errors. The justification for this decision was that a self-correction is more of an error than a repetition because it at least includes a word read incorrectly, whereas a repetition

is merely correct words read multiple times. The complete list of student reading behaviors that were counted as errors in this study are: mispronunciation, substitution, tester provided (refusal to pronounce), omission, insertion, reversal, and self-correction (see Appendix I for a copy of the scoring manual for running records).

For each text, I listened to the audio file and marked oral reading errors directly onto the score sheet, using standard running record markings (see Burns & Roe, 1999, p. 16). I listened to sections of the audio file multiple times as necessary in order to create an accurate record of each student's oral reading performance on a text. I then tallied the different error types on the score sheet and entered the totals into a database. This information was used to calculate the oral reading accuracy rate, by subtracting the number of errors from the total number of words in the passage, and then dividing by the number of words in the passage, yielding a percentage accuracy rate. Although I did the primary scoring on all running records, another graduate student also scored a sample of them to establish interrater reliability, as described in the Data Analysis section below.

*Comprehension Questions.* Using the audiotapes, all questions and responses were transcribed into a database. Questions were then separated from answers so that they could be screened and evaluated based on their quality without being influenced by students' responses. As mentioned earlier, researchers asked an average of 7 questions in the hopes of generating 6 usable ones. In evaluating the questions that had been asked, the first step was to make sure questions were grouped appropriately, so that any follow-up prompts were included with the original question rather than being counted as separate questions. Valmont's (1972) criteria were then used to eliminate any blatantly faulty questions. For example, the question "What do the angels look like?" for the book *Star*

*Wars Episode I Journal: Anakin Skywalker* (Strasser, 1999), was eliminated because it was answerable both from the text and from the picture that accompanied it. The question “Is the little girl afraid of the thunder?” for the book *Thunder Cake* (Polacco, 1990) was eliminated because it offered a fifty-fifty chance of a correct response. After faulty questions like these were discarded, the remaining questions were categorized according to three different facets: 1) what information was being asked for: main idea, detail, sequence, cause and effect, or vocabulary; 2) where the information was located: retelling in fact or putting information together (adapted from Woods & Moe, 2007); and 3) the type of thought processes required: literal recall or inferential thinking. These various categorizations were drawn from a number of different sources (e.g., Applegate et al., 2008; Burns & Roe, 1999; Caldwell, 2002; Johns, 1997; Stieglitz, 2002; Valmont, 1972) and were used to ensure that the questions used for this study were at least representative of the types of questions used in published instruments.

Once this categorization was complete and all questions were labeled, the entire set of questions for each text was examined to see if there was a balance of question types. If there were more than 6 questions remaining for any text, questions were deleted at random from any over-represented categories until a set of 6 questions had been created. Most of the question sets (78.1%) consisted of 6 questions, although some of the shorter or simpler texts only allowed for 5 questions (15.6%) or even 4 questions (6.3%).

When this screening and selection process was complete, the questions were reunited with their corresponding student responses. Each response was then scored on a scale of 0-2, with 0 being entirely incorrect, 1 being partially correct, and 2 being entirely correct. A total score was calculated for each text, and this total score was then converted

to a percentage score by dividing by the total number of possible points, which was usually 12, but sometimes 10 or 8.

*Reading logs.* In entering students' handwritten reading log entries into a database, the relatively large number of entries that were either illegible or incomplete meant that several important decisions had to be made. First, some log entries only listed a book title, with no additional information for any of the short response or scale items. These entries were not included in any of the data analyses, and they were not included in the database. Second, there were a number of entries that included a title and some additional information but contained one or more missing or illegible fields. As a general rule, entries with missing or illegible fields were included in the database but were only used in analyses for which they offered complete information. For example, some students used the scale item to rate the difficulty of their texts but did not indicate their enjoyment by circling from one to five stars. These entries were included in analysis of perceptions of difficulty but not in analysis of text enjoyment. As another example, a small number of entries contained complete information on the "Books I Chose" side of the log but no information on the "Books I Read" side of the log, possibly because students chose texts but later abandoned them before finishing. These entries were used for analysis of students' reasons or purposes for choosing texts, but not for any analysis related to enjoyment or perceptions of difficulty. The same treatment was given to two entries in which students explicitly mentioned abandonment of a text by noting, "I read half of the book" and "I did not choose this book." However, it also seems possible that students sometimes completed full log entries even for abandoned texts; this occurrence is impossible to detect, but it must be acknowledged that the data on perceptions of

difficulty and enjoyment may reflect both the texts that students chose and read and the texts they chose but later abandoned.

### *Interrater Reliability*

*Oral retellings.* I did an initial scoring of all of the retellings and then trained another graduate student in the use of the rubric. After practicing scoring a few retellings together, she then scored five more on her own. When we were confident that we were consistent in our interpretations of the five scoring categories, I gave her 20 transcribed retellings to score on her own. These 20 retellings were drawn as a random, stratified sample across the conditions of easier and more difficult texts. For each of the 20 oral retellings in the interrater sample, I compared my original score with her score. I then calculated the percentage of agreement between our scores, finding that we had exact matches on 15 of the 20 retellings (75%), and we were within 1 point on an additional 4 retellings (20%), for a total score correlation of 0.988. The largest difference between our two scores was 2 points, on a 10 point scale.

*Running records.* I selected a sample of 24 running records, drawing 12 at random from the set of easier texts and 12 at random from the set of more difficult texts, and balancing relatively equally across classrooms. I gave the audio files and text samples to the other graduate student, and she scored them and returned them to me. A comparison of my scores with her scores revealed a correlation of 0.984 at the level of total errors per reading.

*Comprehension questions.* As with the oral retellings and running records, a subset of the comprehension questions and responses was selected for the purposes of establishing interrater reliability. A stratified, random sample of 20 question sets,



balanced according to text difficulty, was given to another graduate student for scoring. Interrater reliability was calculated for the total score rather than for individual items or types of items. This method was deemed appropriate because analysis was conducted at the level of total scores and not at the level of individual questions. We had exact matches on 12 of the 20 sets of comprehension questions (60.0%), and we were within 1 point on an additional 3 sets (15.0%), for a total score correlation of 0.919. The largest score difference was 3 points (on a 12 point scale), and this difference occurred once.

*Interview responses.* To allow for analysis of students' reasons for choosing frustration-level texts, responses to the interview question, "Why did you choose this book/magazine/other?" were coded and categorized for the 35 frustration-level texts and for their 35 independent-level matches. Interrater reliability for this coding system was established by having another graduate student code interview responses for a random sample of 20 interview transcripts. To assist in this process, a coding manual was developed that provided names, descriptions, and examples for each of the coding categories, and I provided the second rater with a brief training session that included joint scoring and discussion of a set of five additional interview transcripts. For the random sample of 20 transcripts, 42 codes were generated, and the second rater agreed with my initial rating for 90.5% of those codes.

#### *Identifying the Subsample of Frustration-Level Choosers*

In order to compare student reading performance on frustration- and independent-level texts, it was necessary to identify a subsample of students such that each had been assessed reading at least one text from each category of text difficulty. One important decision in the identification of this subsample of frustration-level choosers was the

decision to use oral reading accuracy as the sole determinant of the frustration level. The main reason for this decision was that, because comprehension was being tested as an outcome variable, it was untenable to also use it as a determinant of text difficulty. In other words, using comprehension as criterion for frustration-level placement would have made the subsequent analysis of relationships between text difficulty and comprehension highly problematic. Other support for this decision comes from the fact that some IRIs call for determination of frustration-level using *either* the word recognition criterion *or* the reading comprehension criterion, but not both (e.g., Burns & Roe, 1999). In fact, Betts himself (1946) originally intended the frustration level to be determined by either the word recognition or the comprehension criterion. Relying only on oral reading accuracy as a determinant of reading level was thus both methodologically necessary and consistent with assessment methods advocated by Betts and others.

The next decision was to determine which numbers to use as thresholds for the frustration, instructional, and independent reading levels. As described in Chapter 2, in the more than 60 years that have passed since Betts (1946) and Kilgallon (1942) first suggested a framework of reading levels, scholars and IRI publishers have put forth a wide variety of criteria, both for word recognition and for reading comprehension (for a chart comparing different leveling criteria, see Appendix J). Choosing which levels to use was therefore no simple matter. The frustration-level cut-off of 90% oral reading accuracy is fairly consistent across sources, so it was chosen as the upper limit for frustration-level texts. Additional support for this decision comes from the work of Davis (1975), whose polygraph studies of a sample of fourth and fifth graders found that Betts' (1946) 10 percent oral reading error rate corresponded well with polygraph measures of

frustration, when repetitions were counted as errors. In this study, repetitions were not counted as errors, and Davis suggests that this scoring method may result in frustration occurring with an error rate of less than 10 percent. Using 90% as an upper limit for frustration-level may thus be a conservative approach to identifying frustration-level texts, but it is consistent with standard practice.

Based on scores from running records of all 159 assessed texts, oral reading accuracy ranged from a low of 65.2% to a high of 100.0%. Applying the 90% oral reading accuracy criterion to the entire sample of students and texts, 48 of the 159 assessed readings fell into the frustration-level category. When duplicate cases – multiple frustration-level readings by the same student – were removed, 36 cases remained. In other words, out of the entire sample of 70 students, 36 students (51.4%) were assessed reading at least one frustration-level text at some point during the data collection period.

For the independent level, the original plan was to use a lower limit of 99% oral reading accuracy, which is consistent with a number of published IRIs (e.g., Applegate et al., 2008; Burns & Roe, 1999; Johns, 1997; Stieglitz, 2002; Woods & Moe, 2007) and with Betts' original framework (1946). However, as several other researchers have found in studying primary grade readers (e.g., Powell, 1970; Powell & Dunkeld, 1971), it quickly became apparent that this standard was simply too high, at least for the sample of second graders in the study. Even without counting repetitions as errors, students' performance on only 7 of the 159 assessed texts (4.4%) met the independent-level criterion of 99% oral reading accuracy. As a result, the decision was made to use a less stringent lower limit for oral reading accuracy on the easier texts. Several sources advocate using 95% (Armbruster, Lehr, & Osborn, 2003; Clay, 1985) oral reading

accuracy as the lower limit for the independent reading level regardless of age group, and Powell (1980) recommended a cutoff of 94% for readers in grades one and two. Suggestions for the instructional level begin at anywhere from 85 to 95% accuracy. For the purposes of this study, which are both to test the utility of commonly-used criteria and to compare students' performance on difficult and easier texts, selection rules were created such that easier texts were considered to be any texts for which the following two conditions were met: 1) a student achieved at least 92% oral reading accuracy on the text sample; and 2) the difference between the student's oral reading performance on the difficult and the easy texts was at least 3%. The 92% threshold was chosen because it is the lower limit of the instructional level according to Lipson & Wixson (2003, based on a review of published IRIs) and Powell (1970, for oral rereading after silent reading, for first and second grade students), and it also represents the approximate midpoint between the lower limits of some versions of the independent and instructional levels (e.g., Armbruster et al., 2003; Clay, 1985). This looser interpretation of the independent level guidelines was necessary, given the extremely small number of readings that met the common 99% accuracy criterion. The 3% buffer zone was intended to decrease the likelihood that the differences in a student's performance on the two assessed texts were due merely to chance. Although it is a relatively small buffer, the reality is that these subtle differences are often used in practice (e.g., Johnson, Kress, & Pikulski, 1987; Stieglitz, 2002), as teachers follow IRI procedures to make determinations of their students' reading levels. In addition, the high rate of interrater reliability for the running records implies only a 1.165% average error rate in scoring, so the 3% buffer should be adequate for limiting the role of scoring error in affecting reading level determinations.

According to these selection rules, one of the 36 frustration-level cases did not have an independent-level match, so it was removed from the list of potential cases for the subsample of frustration-level choosers, resulting in a final grouping of 35 matched pairs of difficult and easier texts. In these matched pairs, the differences between any individual student's oral reading accuracy on a difficult text and an easier text ranged from 3.7% to 32.8%, with a mean difference of 11.9%. Table 3.3 provides additional information about the distribution of oral reading accuracy scores for the subsample of students who chose frustration-level texts, including mean scores for the groups of frustration- and independent-level texts at each interval of the range of differences.

Table 3.3. Oral reading accuracy differences and mean scores for frustration-level and independent-level texts chosen by the subsample of students who chose to read frustration-level texts.

Difference <sup>a</sup>	n	% of cases	Mean oral reading accuracy	
			Frustration	Independent
3-5%	1	2.9%	90.3%	94.0%
5-10%	17	48.6%	88.3%	96.2%
10-15%	8	22.9%	83.4%	95.7%
15-20%	7	20.0%	77.4%	95.2%
> 20%	2	5.7%	68.2%	95.6%
Total	35	100%	83.9%	95.8%

<sup>a</sup> This column contains information about differences between individual students' oral reading accuracy scores on their frustration- and independent-level texts, which were calculated by subtracting an individual's oral reading accuracy percentage for their frustration-level text from the same measure for their independent-level text. For example, for the interval that includes score differences between 5 and 10 percent, 17 students (48.6% of the subsample of frustration-level choosers) had score differences within this range.

## *Analysis*

Several initial analyses were conducted to gather some general data about the entire sample of 70 students and the texts they chose. In all cases, quantitative data analysis was conducted using the Statistical Package for the Social Sciences software (SPSS, version 16.0). A  $p < 0.05$  level of statistical significance was used. First, descriptive statistics (i.e., minimum, maximum, mean, median, and standard deviation) were generated for GMRT-4 percentile scores and MRP raw scores, both for the subtests and subscales and for the combined scores. Second, to determine relationships between reading ability and motivation to read, Pearson correlations were used to measure correlations between the GMRT-4 Word Decoding subtest, Comprehension subtest, and total score; and the MRP Self-Concept subscale, Value of Reading subscale, and combined score. For these correlations, scatterplots were examined to protect against the possibility that any outliers would disproportionately affect the correlation. Missing data was handled through casewise deletion. Third, relationships between reading ability and gender and between motivation to read and gender were examined through the use of independent samples t-tests. For each t-test, Levene's test was used to see if the data met the assumption of equality of variances, such that the variation of scores for the two gender groups was not significantly different. This assumption was met in each case.

Following these initial analyses, a number of additional procedures were performed on the collected data to address the five research questions. The following sections describe each of these analyses in turn.

*Question 1: Do the students who choose texts at their frustration level fit into certain profiles based on gender, motivation to read, or reading ability?* For this research

question, the first step in data analysis was to use oral reading accuracy scores to identify a subsample of students who were assessed reading both a frustration-level text and an independent-level text. (The process for identifying this subsample was described in the previous section of this chapter.) After identifying the subsample of frustration-level choosers, several different quantitative analyses were used to determine relationships between the tendency to select frustration-level texts and the student variables of reading ability, motivation to read, and gender. Paired samples t-tests were used to test the significance of the differences between students' GMRT-4 percentile scores on the Word Decoding and Comprehension subtests, both for the 35 students who were included in the subsample of frustration-level choosers and for the 35 students who were excluded from it.<sup>3</sup> This same method was used to test differences between MRP scores on the Self-Concept and Value of Reading subscales. Independent samples t-tests were then used to test the significance of the differences between mean scores for the two subsamples of students on the GMRT-4 subtests and on the MRP subscales and combined scores. For each of these t-tests, Levene's test was used to see if the data met the assumption of equality of variances. The assumption of normality was tested by examining histograms to see if variables were normally distributed for both subsamples and by calculating skewness and kurtosis statistics for each distribution. Finally, a Chi-square analysis was used to compare the gender proportions in the subsample of frustration-level choosers to

---

<sup>3</sup> For simplicity's sake, for the remainder of this dissertation I will refer to these two subsamples of students as those who did and did not choose frustration-level texts. This is not meant to imply that students in the first group always (or even often) chose to read frustration-level texts; it merely means that they were identified as having chosen to read at least one text during the data collection period that was determined to be at a frustration-level for them in terms of oral reading accuracy. Similarly, the label for the second group is not intended to suggest that the students in that subsample never chose to read frustration-level texts; they simply were not identified as having done so during the data collection period.

the gender proportions in the sample as a whole. Specifically, the Pearson chi-square test of goodness of fit was used to test the null hypothesis that the percentage of boys in the group of frustration-level choosers was equal to .514, which was the proportion of boys in the entire sample. This test was chosen because the analysis involved nominal data and a relatively large sample ( $n = 70$ ).

*Question 2: What reasons or purposes do students give for choosing frustration-level texts for independent reading?* Analysis of students' reasons or purposes began with a review of all reading log entries from the entire sample. This review was used to get a general sense of the range of explanations given by all students for all texts and to help generate an initial set of coding categories. A similar review was then conducted for interview responses for the subsample of frustration-level choosers, in which the coding categories were applied to students' interview responses for the set of 70 identified texts (35 matched pairs). Frequencies were calculated, allowing for a comparison of students' reasons for choosing frustration-level texts and independent-level texts. Finally, I looked more closely at interview responses that explicitly mentioned text difficulty as a reason or purpose for choosing either a difficult or an easier text.

*Question 3: What, if anything, do students understand from reading self-selected, frustration-level texts (as compared to independent-level texts) independently?* To address this third question, analysis began with an exploration of the relationships between oral reading accuracy (which was used to determine frustration and independent levels) and comprehension. Pearson correlations were used to test the significance of the relationship between oral reading accuracy scores and comprehension question scores, both for the entire group of 70 texts chosen by the subsample of frustration-level choosers



and for the component subgroups of 35 frustration-level and 35 independent-level texts. For these correlations, the unit of analysis was the text, rather than the individual student, and the analysis only included the 70 texts chosen by the subsample of frustration-level choosers. The reason for this decision was that the larger set of 159 assessed texts was not distributed equally across the 70 students in the entire sample. As a result, although the set of 159 reading performances offered a larger dataset, using it would have meant that the reading behaviors of individual students would factor disproportionately into what was meant to be a general relationship between two performance variables. Independent samples t-tests were also used to test the significance of the mean differences of the comprehension question scores between the groups of frustration- and independent-level texts. Levene's test was used to test the assumption of equality of variances; histograms and skewness and kurtosis statistics were used to test the assumption of normality. These t-tests also used individual texts, rather than students, as the unit of analysis. Finally, an informal matched pairs comparison used individual students from the subsample of frustration-level choosers as the unit of analysis to determine whether there were any cases in which a student's comprehension score for the frustration-level text was actually higher than his score for the matched, independent-level text.

*Question 4: What are students' perceptions of the difficulty of their chosen, frustration-level texts?* Analysis related to this question began with a review of reading log entries, examining the distribution of students' ratings of text difficulty by calculating frequencies for each point on the rating scale. Next, a similar review was conducted for texts chosen by students in the subsample of frustration-level choosers, using student

scale responses to the interview question, “How easy or how difficult was this book/magazine/other for you?” The frequencies calculated in this second review allowed for a comparison of ratings for frustration- and independent-level texts. Third, student responses to the interview questions, “Were there words that you didn’t know?” and “Were there parts of the book/magazine/other that you didn’t understand?” were used to explore the possibility of differential roles for oral reading and comprehension in the formulation of perceptions of text difficulty. To this end, response frequencies were calculated and graphed separately for the sets of frustration- and independent-level texts. Finally, the possible role of prior knowledge in perceptions of text difficulty was examined by determining the distribution of difficulty ratings for frustration-level texts according to number of previous readings of a text. Qualitative data were also used as a source for illustrative examples regarding this distribution. Because the analyses related to this fourth research question were intended to be primarily descriptive, statistical significance was not determined for any of the calculated frequencies.

*Question 5: What are students’ affective experiences with reading these difficult texts independently? In other words, are self-selected, frustration-level texts actually “frustrating,” or are students able to enjoy them?* To address this final research question, frequency calculations for reading log entries were used to determine the overall distribution of students’ ratings of enjoyment for the entire sample of selected texts. To examine the possible role of text difficulty in student enjoyment, a comparison of frustration-level choosers’ responses to the interview question, “Did you enjoy reading this book/magazine/other?” was conducted for the sets of frustration- and independent-level texts. Frequencies were calculated, and interview data were used as a source of

illustrative examples. Finally, the statistical significance of the relationship between perceptions of text difficulty and ratings of enjoyment was tested by using a Spearman rank-order correlation for the ordinal data from scale responses to relevant interview items.

## CHAPTER 4: RESULTS

In presenting the results of this study, I begin by relating some general, descriptive results that help to give a picture of the findings as a whole. I then move on to consider results related to each of the five research questions in turn. For the most part, I have reserved discussion and interpretation of the results for the subsequent chapter.

### General Results

#### *Reading Ability*

Gates-MacGinitie Reading Test (GMRT-4) scores were available for 69 of the 70 students in the sample. Descriptive statistics for the Word Decoding subtest, Comprehension subtest, and combined scores of the entire sample are presented in Table 4.1 below, both as raw scores and as percentile ranks.

As mentioned in the Methods chapter, the Word Decoding subtest had 43 items and the Comprehension subtest had 39 items, with a maximum total raw score of 82. For both subtests and for the combined score, the raw scores of the sample show a negative skew, with median scores higher than the mean for all three scores: Word Decoding ( $M = 32.90$ , Median = 36,  $SD = 9.17$ ), Comprehension ( $M = 30.01$ , Median = 32,  $SD = 7.75$ ), and the combined score ( $M = 62.91$ , Median = 66,  $SD = 15.40$ ). This apparent deviation from the normal distribution is not a concern, however, because the GMRT-4 is a norm-referenced test, with percentile scores based on a national norming sample. It can be assumed that GMRT-4 scores will fall along a normal distribution, although in this sample the curve appears to be centered around the 37<sup>th</sup> and 36<sup>th</sup> percentiles rather than the 50<sup>th</sup>. There was some diversity across classrooms, with the mean percentile scores for

each classroom ranging from a low of 39.8 to a high of 65.0 for Word Decoding, and from a low of 33.1 to a high of 60.3 for Comprehension.

Table 4.1. Reading ability scores for the whole sample, from the GMRT-4.

GMRT Subtest <sup>a</sup>	Min	Max	M	Mdn	SD
Word Decoding					
Raw	12	43	32.90	36	9.17
Percentile <sup>b</sup>	2	94	37	47	-
Comprehension					
Raw	6	39	30.01	32	7.75
Percentile <sup>b</sup>	1	98	36	46	-
Combined <sup>c</sup>					
Raw	24	82	62.91	66	15.40

<sup>a</sup> n = 69 for all measures. <sup>b</sup> The published percentile scores for forms S and T of the GMRT-4 differ slightly by form. Because 62 of the 69 tested students (89.9%) took form S, I have chosen to report the percentile scores for that form only. <sup>c</sup> The national percentile ranks provided by the publisher of the GMRT-4 are calculated for the total of all three available subtests. Because this study only used two of the three subtests, percentile scores are unavailable for the combined score.

### *Motivation to Read*

Motivation to Read Profile (MRP) data were available for 66 of the 70 students in the sample. However, because one student left an item on the Value of Reading subscale blank, the sample size for that subscale and for the combined score is only 65. The means and standard deviations for the Self-Concept subscale, Value of Reading subscale, and combined scores of the entire sample are presented in Table 4.2. As described in the Methods chapter, each of the subscales consists of 10 items, which are all scored on a 4-

point ordinal scale. For each subscale, then, possible subscale scores range from 10 to 40, and possible combined scores range from 20 to 80.

Table 4.2. Motivation to read for the whole sample, from the MRP.

MRP Score	n	Min.	Max.	M	Mdn	SD
Self-Concept	66	18	40	31.21	31	5.57
Value of Reading	65	10	40	32.97	35	6.81
Combined	65	32	40	64.28	66	11.13

For the Self-Concept subscale, the scores fall along a fairly normal distribution, with mean and median scores virtually identical ( $M = 31.21$ , Median = 31,  $SD = 5.57$ ). For the Value of Reading subscale and for the combined score, the scores of the sample exhibit a negative skew. As with the reading ability scores, these motivation to read scores are clustered toward the high end of the distribution, as seen in Figure 4.1 below. The median scores are again higher than the mean scores for the Value of Reading subscale ( $M = 32.97$ , Median = 35,  $SD = 6.81$ ) and for the combined score ( $M = 64.28$ , Median = 66,  $SD = 11.13$ ).

The mean scores for both MRP subscales suggest that scores for individual items hovered right around the third point on a four-point scale. To give a better sense of what these scores represent, the third point on the scale for the item “I am \_\_\_\_\_” is “a good reader,” midway between “an OK reader” and “a very good reader.” And for the item “Knowing how to read well is \_\_\_\_\_,” the third point on the scale is “important,” between “sort of important” and “very important.”

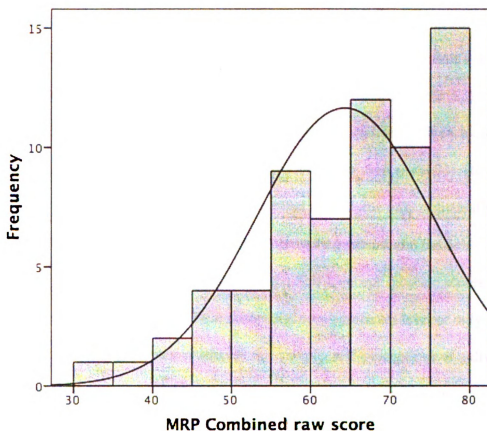


Figure 4.1. Histogram showing distribution of MRP combined raw scores.

*Interactions: Reading Ability, Motivation to Read, and Gender*

Having considered two of the student variables in turn, it is also important to point out some interactions among reading ability, motivation to read, and a third student variable under investigation in this study: gender. This part of the analysis included comparisons of each possible pair of variables: 1) reading ability and motivation to read; 2) reading ability and gender; and 3) motivation to read and gender. A description of the analysis related to each of these three comparisons follows.

*Reading Ability and Motivation to Read.* Previous studies have established positive relationships between motivation and reading achievement, although the directionality and exact nature of that relationship is up for debate, as is its applicability

to readers of different ages (e.g., Guthrie, Wigfield, Metsala, & Cox, 1999; Schunk & Zimmerman, 1997). To test the connection between reading ability and motivation to read for the second graders in this sample, correlations were run between GMRT-4 subtests and total scores and MRP subscales and combined scores (see Table 4.3). As might be expected, the GMRT-4 subtests were significantly correlated with each other ( $r = .654, p < 0.01$ ), as were the subscales of the MRP survey ( $r = .613, p < 0.01$ ). However, between the measures of reading ability and motivation to read, the only statistically significant correlation appeared between the Word Decoding subtest and the Self-Concept subscale ( $r = .326, p < 0.01$ ). In contrast, the Self-Concept subscale scores correlated with the Comprehension subtest scores at a non-significant level of only .078. In other words, for the second grade readers in this sample, self-perceptions of individual reading ability are more closely related to word decoding performance than to comprehension performance.

Table 4.3. Pearson correlations between GMRT-4 and MRP subtests and totals.

Variable	1	2	3	4	5	6
1. Word Decoding		.654**	.925**	.326**	.003	.174
	-	.000	.000	.008	.982	.170
2. Comprehension			.893**	.078	-.108	-.026
		-	.000	.537	.396	.839
3. GMRT-4 total				.229	-.054	.087
			-	.066	.671	.495
4. Self-Concept					.613**	.875**
				-	.000	.000
5. Value of Reading						.919**
					-	.000
6. MRP total	-					
						-

\*\*  $p < 0.01$ (two-tailed)



*Reading Ability and Gender.* The mean raw scores on the GMRT-4 did not differ by gender at a level of statistical significance (see Table 4.4). Boys ( $M = 34.00$ ,  $SD = 7.93$ ) outscored girls ( $M = 31.70$ ,  $SD = 10.34$ ) slightly on the Word Decoding subtest by a mean difference of 2.303, while girls ( $M = 30.39$ ,  $SD = 9.00$ ) outscored boys ( $M = 29.67$ ,  $SD = 6.52$ ) on the Comprehension subtest by a mean difference of 0.727. When the two subtests are combined, the mean raw score for boys ( $M = 63.67$ ,  $SD = 13.15$ ) was higher than that for girls ( $M = 62.09$ ,  $SD = 17.71$ ) by 1.576 points. However, independent samples t-tests comparing the group means on each of these variables revealed no statistically significant gender differences in performance on the GMRT-4, either for the subtests or for the combined scores. The results of this analysis remained the same when using percentile scores rather than raw scores.

Table 4.4. T-tests of reading ability scores, by gender.

GMRT-4 Score	Boys (n=36)		Girls (n=33)		Mean difference	t value (p value)
	M	SD	M	SD		
Word Decoding	34.00	7.93	31.70	10.34	2.303	1.043 (.301)
Comprehension	29.67	6.52	30.39	9.00	-0.727	-.387 (.700)
Combined	63.67	13.15	62.09	17.71	1.576	.422 (.674)

*Motivation to Read and Gender.* Scores on the Motivation to Read Profile did differ between boys and girls at a level of statistical significance. On the Self-Concept subscale, the mean raw score for girls ( $M = 32.22$ ,  $SD = 6.06$ ) was higher than the mean raw score for boys ( $M = 30.26$ ,  $SD = 4.98$ ) by 1.954 points. Girls ( $M = 34.91$ ,  $SD = 5.11$ )

also outscored boys ( $M = 31.09$ ,  $SD = 7.76$ ) on the Value of Reading subscale, with a mean difference of 3.815. Combining the two subscales reveals a mean difference of 5.610, with the mean score for girls ( $M = 67.12$ ,  $SD = 9.98$ ) higher than that for boys ( $M = 61.52$ ,  $SD = 11.63$ ). Independent samples t-tests comparing the means on each of these variables (see Table 4.5) show statistically significant gender differences both on the Value of Reading subscale ( $t(63) = -2.234$ ,  $p < 0.05$ ) and on the combined MRP score ( $t(63) = -2.084$ ,  $p < 0.05$ ), but not on the Self-Concept subscale ( $t(64) = -1.435$ ,  $p = 0.156$ ).

Table 4.5. T-tests of motivation to read scores, by gender.

MRP Score	Boys <sup>a</sup>		Girls (n = 32)		MD	t value (p value)
	M	SD	M	SD		
Self-Concept	30.26	4.98	32.22	6.06	-1.954	-1.435 (.156)
Value of Reading	31.09	7.76	34.91	5.11	-3.815	-2.334 (.023)
Combined	61.52	11.63	67.12	9.98	-5.610	-2.084 (.041)

<sup>a</sup> For the Self-Concept subscale,  $n = 34$ ; for the Value of Reading subscale and the Combined score,  $n = 33$ .

### *Texts*

Across the 70 students in the 5 participating classrooms, one-on-one assessments were conducted for a total of 159 texts, for an average of 2.27 texts assessed per student. The actual number of texts assessed per student ranged from 1 to 4, with most students assessed for two different texts. As mentioned in the Methods chapter, several students were assessed on more than two different texts because they had already read a

frustration-level text, but it was difficult to find an independent-level text to create a matched pair. The 159 assessed texts represented a wide range in terms of genre, length, and readability. Fiction choices included narrative picture books, easy readers, chapter books, graphic novels, and comic books; non-fiction choices included magazines, informational texts, and even a video game instruction manual. Still other texts, like poetry collections and hybrid narrative/expository texts defy simple classification.

To gain information about text readability, Lexile scores were determined for each of the 159 assessed texts. As explained in Chapter 3, the Lexile framework is a quantitative method of measuring both text readability and individual reading ability on a single scale. The Lexile website offers a large database of popular texts and an online text analyzer tool, both of which were used to obtain Lexile scores for the assessed texts. Lexile scores for these 159 texts ranged from a low of BR – which stands for “Beginning Reader” and is used for any texts with a Lexile measure of zero or below – to a high of 1460L. Texts at the low end of this range include the rhyming book *Hop on Pop* (Seuss, 1963) and the leveled readers *Hide and Seek* (Brown & Carey, 1994) and *Who Lives in the Rainforest?* (Canizares & Reed, 1998). Texts at the high end of the range include the picture book *Eat Your Peas* (Gray, 2000) and a number of informational texts, such as *Dragonology: The Complete Book of Dragons* (Drake & Steer, 2003), *The Mind of the Cat* (Brotsky, 1990), *Ghost Liners: Exploring the World’s Greatest Lost Ships* (Ballard & Archbold, 1998), and *Young People’s Atlas of the United States* (Harrison, 1992). The mean Lexile score across the sample of 159 texts was 505L, which is roughly equivalent to a 3rd grade level. Table 4.6 below shows the distribution of assessed texts across the range of Lexile scores and approximate grade level equivalents (GLEs).

It is important to note that the sample of texts described above is not necessarily representative of the sample of texts chosen by students as a whole. The assessed texts were selected purposively, rather than randomly, from the total set of students' text choices during the data collection period. This purposive sampling procedure – conducted with the express intention of identifying texts that I expected to be particularly difficult or particularly easy for individual students, based on their reading ability scores – understandably impacts the readability data for the sample of assessed texts. As a result, this data cannot be used as the basis for any conclusions about students' preferences for texts written at certain readability levels. It is, however, important contextual information that will aid in the interpretation of later results.

Table 4.6. Assessed texts by Lexile level and approximate grade level equivalent.

Lexile Range	GLE	n	% of Texts	Lexile Mean
< 200L	K-1	20	12.6%	73L
200-350L	1	30	18.9%	301L
350-500L	2	42	26.4%	442L
500-650L	3	30	18.9%	569L
650-850L	4	16	10.1%	731L
> 850L	5+	21	13.2%	1070L
TOTAL	-	159	100.0%	505L

I now turn to reporting the results for each of the five research questions. A discussion of findings integrated across questions is found in Chapter 5.

Question 1: Do the students who choose texts at their frustration level fit into certain profiles based on gender, motivation to read, or reading ability?

The previous part of this chapter described some characteristics of the students and texts in the sample as a whole. In addition, it pointed out some potentially interesting interactions among the three student variables that are the focus of this first research question. In addressing the question, the analysis now moves from a general look at the entire sample toward a closer examination of the subsample of students who chose to read frustration-level texts at some point during the data collection period.

Having identified 35 occasions in which students chose to read a frustration-level text, I addressed the first research question by describing the students included in the subsample of frustration-level choosers and then comparing them to the subsample of students who did not choose to read frustration-level texts. This description and comparison was conducted for the three student variables of reading ability, motivation to read, and gender. In addition to considering these three student variables individually, I also looked for interactions among them.

### *Reading Ability*

For the subsample of 35 students who chose to read frustration-level texts, GMRT-4 scores were available for 34 of them. For this subsample, percentile scores on the Comprehension subtest ( $M = 37.47$ ,  $SD = 29.15$ ) were slightly higher than percentile scores on the Word Decoding subtest ( $M = 30.35$ ,  $SD = 25.62$ ). A paired samples t-test showed that this observed difference was not statistically significant,  $t(33) = -1.896$ ,  $p = .067$  (two-tailed). In contrast, the group of students who did not choose frustration level

texts ( $n = 35$ ) had slightly higher mean scores on the Word Decoding subtest ( $M = 63.71$ ,  $SD = 24.77$ ) than on the Comprehension subtest ( $M = 60.40$ ,  $SD = 26.31$ ), although a paired samples t-test showed that this mean difference was also non-significant,  $t(34) = 0.816$ ,  $p = .420$  (two-tailed).

To further compare the reading ability of frustration-level choosers with non-choosers, independent sample t-tests were used to compare means for the percentile scores of the two groups on the GMRT-4 subtests. As Table 4.7 shows, these t-tests revealed significant differences in reading ability between the two groups for both subtests. The group of non-choosers outperformed the subsample of students who chose frustration-level texts on the Word Decoding subtest,  $t(67) = -5.500$ ,  $p = .000$  and on the Comprehension subtest,  $t(67) = -5.000$ ,  $p = .001$ .

Table 4.7. T-tests of reading ability percentile scores, by subsample membership.

GMRT-4 Subtest	Frustration Level Choosers ( $n=34$ )		Non-frustration Choosers ( $n=35$ )		MD	t value (p value)
	M	SD	M	SD		
Word Decoding	30.35	25.62	63.71	24.77	-33.361	-5.500 (.000)**
Comprehension	37.47	29.15	60.40	26.31	-4.447	-5.000 (.001)**

\*\* $p < 0.01$

On average, then, students who chose to read frustration-level texts demonstrated slightly higher comprehension scores than word decoding scores. In comparison to their peers, students who chose frustration-level texts had significantly lower word decoding and comprehension scores than students who did not choose frustration-level texts.

### *Motivation to Read*

For the 35 students who chose to read frustration-level texts, MRP survey scores were available for 31 of them. Paired sample t-tests showed that there were no statistically significant differences between the frustration level choosers' scores on the Self-Concept subscale ( $M = 31.26$ ,  $SD = 5.75$ ) and the Value of Reading subscale ( $M = 33.16$ ,  $SD = 7.26$ ),  $t(30) = -1.663$ ,  $p = .107$  (two-tailed). Similarly, the group of students who did not choose frustration-level texts had only a small, statistically insignificant mean difference between their Self-Concept scores ( $M = 31.35$ ,  $SD = 5.47$ ) and their Value of Reading scores ( $M = 32.79$ ,  $SD = 6.49$ ),  $t(33) = -1.755$ ,  $p = .088$  (two-tailed).

As with the GMRT-4 reading ability scores, independent samples t-tests were used to determine whether the MRP subscale and combined means for the subsample of frustration-level choosers differed significantly from those of the students who were not assessed reading frustration-level texts. These t-tests showed no significant differences between the means of the two populations, either for the Self-Concept subscale,  $t(64) = -.063$ ,  $p = .950$ ; the Value of Reading subscale,  $t(63) = .215$ ,  $p = .830$ ; or the combined MRP score,  $t(63) = .098$ ,  $p = .922$ . Because the MRP is not norm-referenced, it is impossible to determine how these scores compare to the scores of typical second grade readers. However, the relatively high mean scores for the sample as a whole and the lack of significant differences between mean scores for the subgroups suggest that that students who chose frustration-level texts generally had positive self-concepts and beliefs about the value of reading, despite the fact that their reading ability scores were significantly lower than those of their peers.

Table 4.8. T-tests of motivation to read scores, by subsample membership.

MRP Score	Frustration Level Choosers (n = 31)		Non-frustration Choosers (n = 34)		MD	t value (p value)
	M	SD	M	SD		
Self-Concept	31.26	5.75	31.35	5.47	-.087	-.063 (.950)
Value of Reading	33.16	7.26	32.79	6.49	.367	.215 (.830)
Combined	64.42	11.44	64.15	11.01	.272	.098 (.922)

### *Gender*

As described in the previous chapter, the entire sample of 70 students was nearly evenly split along gender lines, with 36 boys (51.4%) and 34 girls (48.6%). In the subsample of 35 students who chose frustration-level texts, however, the distribution of gender was quite a bit different, with 13 boys (37.1%) and 22 girls (62.9%) assessed reading frustration-level texts at some point during the data collection period. In comparison, the group of students who did not choose frustration-level texts was composed of 23 boys (65.7%) and 12 girls (34.3%). Figure 4.2 shows the gender breakdown for the subsample of students who chose frustration-level texts (n=35), the group of students who did not choose frustration-level texts (n=35), and the total sample (n=70).



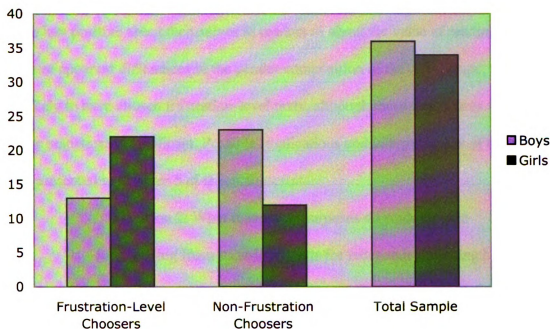


Figure 4.2. Gender breakdown of the total sample and of the two subsamples: frustration-level choosers and non-frustration choosers.

However, despite the apparent differences in gender composition for the subsample and the excluded students, Chi-square analyses reveal no statistically significant differences between the gender composition of either of these subgroups and the sample as a whole. For the subsample of frustration-level choosers,  $\chi^2(1, N = 35) = 2.848, p = .091$ ; for the group of non-choosers,  $\chi^2(1, N = 35) = 2.871, p = .090$ . In other words, the gender proportions in both subgroups were not statistically significantly different from the gender proportion in the total sample.

#### *Question 1: Summary*

Question 1 focused on individual differences among readers who chose to read frustration-level texts. In terms of reading ability, perhaps not surprisingly, students who were in the subsample who chose frustration-level texts were more likely to be lower-

skilled readers than were their counterparts in the group of non-choosers, although they were no less positive about themselves as readers or about the value of reading. Also, there were apparent differences in the gender composition of the two groups, with girls more likely to have chosen frustration-level texts, although these differences turned out not to be statistically significant. Possible implications and interpretations of these results will be presented in Chapter 5 of this manuscript.

Question 2: What reasons or purposes do students give  
for choosing frustration-level texts for independent reading?

In many ways, this second question is an essential counterpart to the first question. Having found out more about which general types of students chose to read difficult texts, an obvious, related issue involves their reasons for doing so. Identifying a phenomenon is not the same as explaining it, so this next section of the results aims to use qualitative data to supplement the general patterns detailed above. To do so, the following analysis relies heavily on data from student reading logs and interview responses. Before moving further, it is important to mention that students' written responses to the reading log prompt "I chose this book because..." tended to be very short and superficial, with numerous, repeated responses like "it looks cool" or "it's a good book" that offered little information about students' thinking. However, there were also quite a few responses that offered more detailed insights into students' reasons or purposes for choosing texts of varying levels of difficulty. The student interview data

typically include more elaborated responses to the question of why they chose their texts, so interview responses provide important additional information.<sup>4</sup>

### *Overview of Reasons for Text Choices*

Across the total sample of 70 students, 1020 reading log entries were collected and recorded, creating a rich source of information related to students' text choices and reading experiences. In their log entries, students gave a wide range of reasons for choosing texts to read. Some expressed general interest in a text's topic or content (e.g., "I like space," "bats are cool," or "it is about sharks"), while others expressed learning goals that were either general (e.g., "I want to learn something" or "I wanted to learn something new") or specific (e.g., "I want to know how to make ice cream," "I want to know how the turtle got his shell," or "I wanted to learn about the first flight").

Some students explained their choices by means of personal connections to the text's content (e.g., "I have chickens," "I have a loose tooth," or "I have a connection because I stay up late"). Others gave social reasons for their choices (e.g., "I wanted to partner read with D [a classmate)]," "M [a classmate] is reading it, and we can talk about it," or "my friend reads it"). Students also frequently expressed preferences for favorite authors (e.g., "I like Patricia Polacco's books" or "I like books written by Robert Munsch); series (e.g., "I like the Boxcar Children series" or "I like Little House on the Prairie"); characters (e.g., "because I like Jack and Annie" or "I like Amelia Bedelia");

---

<sup>4</sup> More generally, due to problems such as incomplete entries, illegible responses, and lost or missing pages, the reading log data were not useful as sources of additional information about individual texts chosen by frustration-level choosers. For this reason, reading log entries were used only in analyses intended to give broad descriptions of the set of selected (and recorded) texts as a whole ( $n = 1020$ ). In contrast, interview data were available for all of the 159 individually assessed texts; these data were therefore used for analyses involving comparisons between the groups of students who did and did not choose to read frustration-level texts.

genres (e.g., “fairy tales are nice,” “I like legends,” or “it’s non-fiction”); or formats (e.g., “I like chapter books” or “I like magazines”).

Some students’ comments were directed toward surface features of their chosen texts (e.g., “it had a rainbow,” “I like the cover, or “I liked the photos”). Others were focused on enjoyment of their texts (e.g., “I think it’s funny,” “I like spooky books,” or “I just wanted to entertain myself”). Students also gave reasons related to the writing style of a text (e.g., “I love the dialogue,” “I love the way it started,” or “it’s got just the right amount of action”) or to their general curiosity about what the text might contain (e.g., “I wonder why the title is called like that,” “I want to know how she swallowed the fly,” or “I want to know the surprise”). And finally, two somewhat opposing reasons for choosing particular texts were related to enjoyment of both familiarity (e.g., “I’ve heard this story a lot of times” or “I read it before”) and novelty (e.g., “I have not read it before,” “I’ve read a lot of books and I’ve not read this book yet,” or “I don’t know that much about wolves”).

Interestingly, one reason for choosing that was rarely mentioned in students’ reading log entries is difficulty level. One girl expressed a preference for reading easier texts, explaining her choice of *The Talented Clementine* (Pennypacker, 2007) by saying, “I read some of it and it’s easy for me to read now.” Later, the same girl described having chosen *Why Do Horses Neigh?* (Holub, 2003) because “it looks easy.” One boy conveyed a desire to avoid too-difficult texts, explaining his choice of an issue of *National Geographic Kids* magazine by saying, “the other book was too hard.” And one student mentioned a preference for more challenging texts (or at least a rejection of too easy

texts), listing *How a Book is Made* (Alik, 1988) in her reading log but then commenting that she did not end up reading the book because it was too easy.

It is important to remember that students wrote these reading log entries before they actually read the texts they were choosing. Their comments related to their purposes for choosing may therefore be constrained somewhat by their limited knowledge of the texts they were about to read. In some cases, students were choosing familiar texts and were in a good position to comment on the text's actual difficulty for them. But in many cases, they could only rely on estimations of difficulty based on what they knew of the text from a cursory review. Still, however, it is interesting to see that even students who ended up rating their chosen texts to be "kind of hard" or "too hard" did not appear to have chosen their texts with difficulty (or ease) in mind. In other words, students' reading log responses do not indicate that students were particularly motivated to seek out or avoid potentially challenging texts. In many cases, their other, varied interests in pursuing texts appear to take primacy over their considerations of text difficulty, although it is impossible to say for sure. The best that can be said from the available data is that students only rarely made explicit mention of text difficulty in their written explanations of their text choices.

#### *Reasons for Choosing Frustration-level Texts*

Having reviewed the larger universe of reasons that students in this sample gave for choosing texts of any level of difficulty, the next analysis focuses more closely on students' reasons for choosing texts that turned out to be frustration-level texts for them in terms of oral reading accuracy. As mentioned earlier, because reading logs were filled out and submitted somewhat inconsistently, information about reasons for choosing a text

were not available for all of the frustration-level texts that were identified. As a result, this analysis involves an examination of the responses that the 35 students in the subsample gave to the first question from the interview protocol: “Why did you choose this book (or magazine, or other format)?” As described in Chapter 3, these interviews took place shortly after a student had finished reading a selected text and directly after completing an oral retelling, a running record, and a set of passage-specific comprehension questions. As with the reading log data, I will present an overview of students’ interview responses and then focus more closely on comments related to text difficulty.

On the whole, both for frustration-level and independent-level texts, the interview responses from the subsample of student who chose frustration-level texts fell into a set of categories highly similar to the ones described in the previous overview of reading log entries. For example, there was a similar focus on reasons related to things like topic interest (e.g., “because I like to learn about dragons” or “because I love horses”), favorite series (e.g., “I like Henry and Mudge books” or “I really like Magic Tree House”), and entertainment (“it’s really funny, and it’s really, really, very, very funny”).

In addition, however, two new categories that had not been present in the reading log data appeared in students’ interview responses. The first of these new categories is classroom connections, which includes any responses that mention classroom instructional activities. For example, one student explained her choice of the book *Animal Poems* (Curry, 2004) by saying,

It's all about poems, and it tells me more about poems 'cause we're working on poems. And if we're writing about poems we need to get a book about poems, like this, and find a page that's important and write about it.

Another student explained his choice of *Franklin in the Dark* (Bourgeois, 1987) by connecting his choice of a book with a turtle as the main character to his recent classroom experiences learning about animals that live in water:

It's kind of interesting to me because we're learning about things that swim in water and I picked the turtle because we have a couple of tadpoles in our room and I'm learning about things that swim in water sometimes, like turtles and stuff.

These two sample responses both fit into the category of classroom connections, although they deal with slightly different purposes for reading. In the first instance, the girl chose the poetry book because she felt it would help her with her poetry writing assignment. In contrast, the boy's choice of a Franklin book seems more like an outgrowth of his developing, situational interest in an instructional topic. In all, 6 of the 35 students who chose frustration-level texts explicitly mentioned classroom connections as a reason for their text choices.

Another potentially new category of purposes for reading involves improving as a reader, although this reason only appeared in the comments of 1 of the 35 students. However, her elaborated response allows us to understand not just that she chose the text because she thinks it will make her a better reader, but that she also has specific ideas about exactly how it will help her improve her reading:

I really like Magic Tree House, and because it has chapters, and sometimes it's a challenge for me to try to read things because I want to be a better reader. So in

kindergarten I had little books and I got higher and higher, and I'm hoping next year that I might be able to read higher books... Some words are very, very hard for me to read and pronounce, and some words, well, like, I don't know a word that my teacher says and this book maybe has the same word, and maybe it will, like, describe it or something.

In other words, she believes that challenging texts will help her improve her reading, in part by exposing her to unknown words in a way that will help her learn what they mean. Interestingly, she was actually able to read the text in question – *Season of the Sandstorms* (Osborne, 2005) – at an independent level.

This overview of interview responses by the subsample of frustration level choosers complements the reading log data by giving more detailed examples of the categories identified earlier and by adding two new categories to the mix. However, another important approach to addressing the second research question lies in finding out whether there were any differences between students' reasons or purposes for choosing frustration-level texts, as compared to their reasons for choosing easier texts. To do this, interview responses for the 35 frustration-level texts and the 35 independent-level matches were coded and categorized.

A comparison of these coded interview responses for independent- and frustration-level texts offers some interesting results, which are summarized in Table 4.9 below. In total, responses fit into one of 22 different categories; for the purposes of this analysis, I have chosen to focus on the response categories that occurred most frequently. As a result, the table only includes those categories that were mentioned in relation to at least 10.0% of either the frustration-level or independent-level cases.



Table 4.9. Most common reasons for choosing frustration- and independent-level texts.

Reason	Frustration		Independent		Total Mentions	
	n	% <sup>a</sup>	n	% <sup>a</sup>	n	% <sup>a</sup>
Classroom	4	12.1%	2	7.7%	6	5.9%
Connection						
Difficulty	1	3.0%	3	11.5%	4	3.9%
Entertainment	8	24.2%	8	30.8%	16	15.7%
Familiarity	8	24.2%	2	7.7%	10	9.8%
General	4	12.1%	1	3.8%	5	4.9%
Information	4	12.1%	3	11.5%	7	6.9%
Personal Connection	4	12.1%	4	15.4%	8	7.8%
Series	0	0.0%	4	15.4%	4	3.9%
Story Interest	5	15.2%	0	0.0%	5	4.9%
Topic Interest	12	36.4%	6	23.1%	18	17.6%

<sup>a</sup> Values in these two columns were calculated by dividing the number of times the category was mentioned by the total number of texts for which this interview response was available. For the group of frustration-level texts,  $n = 33$ ; for independent-level texts,  $n = 26$ . These numbers are less than 35 because the question was not asked for researcher-provided texts. <sup>b</sup> Values in this column were calculated by dividing the number of times a category was mentioned by the total number of category codes generated for the sample of 59 texts ( $n = 102$ ).

Among frustration-level texts, the most commonly cited reason for having chosen a particular text was general interest in the topic; this reason was mentioned in reference to 36.4% of the frustration-level texts in the subsample. The other categories that received mention for at least 10% of the assessed texts were familiarity (24.2% of texts), entertainment (24.2%), story interest (15.2%), personal connection (12.1%), classroom

connection (12.1%), information (12.1%), and general (12.1%). Examples of the types of responses that fit into each of these coding categories can be found in Appendix K. In contrast, for independent-level texts, the most commonly cited reason for choosing a particular text was entertainment, which was mentioned in reference to 30.8% of the interviewed, independent-level texts. The other 5 categories mentioned in regard to at least 10% of the interviewed, independent-level texts were topic interest (23.1% of texts), favorite series (15.4%), personal connection (15.4%), desire for information (11.5%), and level of difficulty (11.5%).

Based on this quantitative summary of the interview data for the subsample of students who chose frustration-level texts, it appears that there may be some differences in the reasons why students chose difficult and easier texts. Particularly striking are the tendencies of students in this group to report having chosen frustration-level texts, but not independent-level texts, for reasons related to topic interest (e.g., “because I really like cats, and I wanted to know what they’re kind of like”), story interest (e.g., “I wanted to see what he would invent in the story”), and familiarity (e.g., “because I seen, like, cartoons of it on Looney Tunes, then I thought of reading it”).

One final source of insights into the question of students’ different purposes for choosing difficult and easier texts comes from the interview responses that deal directly with text difficulty in some way. For the subsample of frustration-level choosers, difficulty was mentioned in relation to only four texts; for the subsample of non-choosers, it was mentioned in regard to five texts. In the paragraphs that follow, some comments related to the role of text difficulty in students’ reading choices for the two subsamples

are integrated and grouped into two categories: those that viewed challenging texts as positive, and those that viewed easier reading as positive.

*Challenge is good.* A few students mentioned potential benefits of reading more challenging texts. As described earlier, one girl in the subsample of frustration-level choosers mentioned having chosen *Season of the Sandstorms* (Osborne, 2005) – a 106-page chapter book – specifically because she wanted a challenge and because she saw the book as an opportunity to improve her own reading by learning the meanings of new words. A girl from the subsample of non-choosers described her choice of *Henry and Mudge: The First Book* (Rylant, 1987) by saying, “Sometimes I really feel like reading long books, and this is like, almost a chapter book, but it just doesn't have chapters.” Second grade is somewhat of a transitional year between picture books or easy readers and longer, chapter books (Bass, 2006), and this particular reader was expressing her desire to become a “chapter book reader.” *Henry and Mudge* had short chapters, so it offered some of the advantages of being a chapter book without being too long or difficult. For both of these cases, the books were actually at an independent level for the girls who chose them; the perceived challenge of the text may have come from the length rather than the actual difficulty level of the text.

*Easy is good.* Other students expressed positive reasons for choosing easier books. Interestingly, while the two students mentioned in the previous paragraph were both girls, the four students in this current category are all boys. One boy, a frustration-level chooser, explained simply that he chose the decodable text *My Life on an Island* (James, 2000) “‘cause it’s kind of easy.” Another boy from the subsample of frustration-level choosers, who earlier had struggled through *Toestomper and the Caterpillars*

(Collicott, 1999) with an oral reading accuracy rate of only 65.5%, subsequently chose to read *Green Eggs and Ham* (Seuss, 1960). As he said, “I thought it was good for me to read. It's a little easier and not that hard.” Similarly, another boy, a non-chooser, described his decision to read an easier book, *A Fly Went By* (McClintock, 1958), almost as a vacation from the more difficult chapter books that he had been in the habit of reading. As he explained,

I went to ‘Number Two’ [the bin of second grade level books], because I knew I could read it. Number two means it's the second grade level. That's easy for me. I didn't want to get another hard one because it takes me about a month to read them. I can read most of the books in there, I just have a hard time reading those ones [points to bin of chapter books]. This is my first time reading it by myself. I don't usually pick number two books; I usually read chapter books.

It is worth noting that this particular student was one of the lowest readers in a class of very good readers. In addition, he had just finished reading a longer, more difficult book (*Diary of a Wimpy Kid: Rodrick Rules* [Kinney, 2008]), which he had chosen because “I started reading the first one, and I started to like the series. A, A, and G [classmates] have it, too.” His reading choices may be a case of a struggling reader who often chooses to read above his level for social reasons but who occasionally enjoys taking a break by reading an easier text.

A fourth boy from the subsample of frustration-level choosers explained his choice of the frustration-level text *Star Wars Episode I Journal: Anakin Skywalker* (Strasser, 1999) by saying:

I read this, and I have all the movies of it, so I just knew what happens and I haven't read this book before. It was easy because I've saw all the movies, and I knew what was going to happen.

This quote reveals a likely connection between familiarity and text difficulty, with the student's prior knowledge of the story effectively making the difficult text "easier" to read. Although the book turned out not to be as easy as he perceived it to be, he clearly chose the text at least partly because he perceived it to be easy and familiar.

#### *Question 2: Summary*

Question 2 centered around students' reasons and purposes for choosing frustration-level texts. To get at this information, I began by discussing the larger set of reasons that the entire sample of students gave for reading texts at any level. This analysis of reading log data provided an initial set of categories that was then used to describe the subsample students' reasons for choosing difficult texts. In comparison to their easier counterparts, difficult texts appear to be chosen more often because of interest in a topic, interest in or curiosity about a story, and familiarity with a text, often in the form of prior experiences with the selected text. Finally, an examination of interview responses related directly to text difficulty showed that students had different reasons for seeking out texts based on the level of challenge they were perceived to offer. However, these reasons were similar across the subsamples of students who did and did not choose frustration-level texts.

Question 3: What, if anything, do students understand from reading self-selected, frustration-level texts (as compared to independent-level texts) independently?

To some extent, this third research question is the real crux of the study. Even if students have good reasons for choosing difficult texts, and even if they report that they enjoy reading them, allowing students to read difficult texts independently in the classroom would still be problematic if it turned out that students do not actually understand the texts they choose to read. To address this question, I looked at relationships between oral reading accuracy and comprehension, and I compared the comprehension of the subsample of frustration-level choosers for the easier and more difficult texts that they chose to read.

#### *Oral Reading Accuracy and Comprehension*

Because oral reading accuracy was the variable by which texts were determined to be either at a frustration level or an independent level for the students who chose them, it was also a logical variable to use when looking at the relationship between comprehension and text difficulty. To measure this relationship, a Pearson correlation was run between the variables of oral reading accuracy and comprehension (based on percentage scores on the sets of passage-specific reading comprehension questions) for the 70 texts chosen by the subsample of frustration-level choosers. The analysis showed a statistically significant correlation of .472 ( $p < .01$ , two-tailed) between oral reading accuracy percentage ( $M = 89.86\%$ ,  $SD = 7.59$ ) and percentage scores on the sets of passage-specific comprehension questions ( $M = 72.84\%$ ,  $SD = 22.13$ ). For the 70 texts in the subsample of matched cases, then, there was a statistically significant relationship between the accuracy with which students read the texts aloud and the degree to which

students were able to answer comprehension questions about what they had just read. However, it is important to note that although this correlation is statistically significant, it is not particularly high. In addition, when Betts' reading level criteria are applied to this set of 70 readings, there is a considerable lack of correspondence between frustration-level and independent-level designations based on oral reading accuracy percentages and performance on comprehension questions. For example, of the 35 frustration-level texts, with oral reading accuracy percentages of 90% or less, only 14 (40.0%) had frustration-level comprehension rates of 50% or less. And of the matching independent-level texts, with oral reading scores of 92% or higher, only 15 (42.9%) had independent-level comprehension rates of 90% or above. In other words, although there was a statistically significant relationship between oral reading accuracy and comprehension in general, this correlation was not particularly strong, and there were a number of cases that did not quite fit the Betts model for both oral reading accuracy and reading comprehension.

It is important to note here that this lack of consistent alignment between oral reading and comprehension complicates some of the upcoming analyses and results somewhat. As mentioned earlier, I decided to base determinations of difficulty level entirely on oral reading accuracy and not on comprehension. While this decision was methodologically necessary and conceptually justified, the weakness of the correlation between word reading and comprehension – although it was significant – means that a “difficult text” as defined by this study may or may not be equally “difficult” in terms of comprehension. In the results still to come, I have tried to address this issue by being consistent in distinguishing between word reading difficulty and comprehension difficulty. However, the fact remains that the oral-reading-only determination I have

chosen to use may still mask some of the variability in student reading performance that is related to the inconsistent relationship between oral reading and comprehension.

#### *Comprehension of Frustration- and Independent-level Texts*

In comparing students' performance on their difficult and easier texts, I began by comparing the mean comprehension scores for the groups of independent-level and frustration-level texts. An independent-samples t-test was run to test the difference between the mean comprehension scores of the group of frustration-level texts and their matched, independent-level counterparts. For the subsample of frustration-level choosers, their mean comprehension score for frustration-level texts ( $M = 60.93$ ,  $SD = 22.20$ ) was significantly lower than the mean comprehension score for independent-level texts ( $M = 85.10$ ,  $SD = 14.01$ ), resulting in a mean difference of  $-24.17$ ,  $t(67) = 5.425$ ,  $p < .01$ .

The next step was to look more closely at each matched pair, this time using the individual student as the unit of analysis. Examining the 35 matched cases as individual pairs, we see that students' comprehension of independent-level texts was higher than their comprehension of frustration-level texts in 25 cases, equal in 2 cases, and lower in 7 cases. In cases in which independent-level comprehension exceeded frustration-level comprehension, the gap between the scores ranged from a low of 8.3% to a high of 75.0%. For the 5 cases in which frustration-level comprehension actually exceeded independent-level comprehension, the gap between the two scores ranged from a low of 8.3% to a high of 20.0%.

In terms of comprehension for the frustration level cases, scores on the sets of comprehension scores ranged quite a bit, from a low of 25.0% to a high of 100.0%. At the lower end of this range, there were often obvious connections between word



recognition errors and comprehension difficulties. One girl read a passage from the book *Lucy on the Loose* (Cooper, 2000) – which tells the story of a dog who runs away when the boy who owns her lets her off the leash – with 78.9% oral reading accuracy and 25.0% comprehension. Her oral reading performance included numerous errors that appeared to affect her comprehension, including misreading the boy’s name “Shawn” as “swim” and misreading the dog’s name “Lucy” as “lucky.” As a result, she was unable to answer questions about the characters in the passage. At the higher end of the comprehension range, a boy read a passage from *The True Story of the Three Little Pigs* (Scieszka, 1996) with only 81.8% oral reading accuracy but with 91.7% comprehension. He made a high number of word recognition errors, including several in the printed sentence, “So of course the minute I knocked on the door, it fell right in,” which he read as “So I cried the mate I knocked on the door, I feeled right in.” Despite these errors, he was somehow still able to give a correct response to the cause and effect question, “What happened when the wolf knocked on the door?” Overall, students’ comprehension of frustration-level texts fell across a wide range, with some scores indicating serious gaps or errors in understanding and others suggesting little or no comprehension difficulty despite frequent word reading errors.

### *Question 3: Summary*

This third question focused on students’ understanding of difficult texts in comparison to their understanding of easier texts. A Pearson correlation revealed a statistically significant overall relationship between oral reading accuracy and comprehension, although the correlation was not particularly high ( $r = 0.472$ ). As a result, it was expected that comprehension scores for independent- and frustration-level texts

would differ, given that the groups had been defined and determined based on oral reading accuracy. Indeed, a t-test showed a statistically significant difference in mean comprehension scores between the two groups of texts. Next, a review of the individual matched pairs showed that comprehension of frustration-level texts was lower than comprehension of their independent-level matches for 73.5% of the subsample students, equal for 5.9% of the students, and higher for 20.6%. Finally, a closer look at students' comprehension of their frustration-level texts shows that comprehension scores actually covered a wide range, including some with 100% comprehension despite an oral reading error rate of more than one out of every ten words.

#### Question 4: What are students' perceptions of the difficulty of their chosen texts?

As mentioned in the discussion of the expectancy x value model in Chapter 2, an individual's perception of a task's difficulty may be more important than the actual difficulty of the task in determining achievement behavior (Eccles, 1983). In this study, it was therefore important to find out not only how difficult specific texts were for specific students, but also how difficult the students perceived them to be. Data related to perceptions of text difficulty was available both from reading log entries and from interview responses.<sup>5</sup> This information about students' perceptions is helpful in two different ways. First, by considering ratings of text difficulty in isolation, we can gather general information about students' perceptions of the texts they chose to read. Second,

---

<sup>5</sup> As a point of clarification, this current analysis of student comments related to text difficulty is intended to illuminate issues related to the accuracy of students' perceptions of a text's difficulty for them. This focus on perceptions of difficulty distinguishes it from the previous analysis of similar student comments, which centered on the role of text difficulty as a reason for choosing to read particular texts. Thus, although the data sources are similar (reading log entries and student interview responses), the focus is different.

by considering their perceptions in relation to their assessed performance on the same texts, we can make some additional determinations about the reliability of student judgments of text difficulty. The following sections offer analyses that address both of these points.

### *Overall Perceptions of Text Difficulty*

The first analysis draws on students' reading logs entries, specifically their post-reading ratings of text difficulty. The written logs prompted students to check a box to indicate their rating of the text's difficulty for them, based on a 5-item scale: too easy, kind of easy, just right, kind of hard, or too hard. From the total sample of 1020 reading log entries, difficulty ratings were available for 988 entries. The distribution of difficulty ratings across those 5 categories of difficulty is presented in Table 4.10 below. The most common ratings were "too easy" (47.3%) and "just right" (32.7%); those two categories together comprised 80.0% of all difficulty ratings. The third most common rating was "kind of easy" (11.9%), followed by "kind of hard" (5.9%) and "too hard" (2.2%). It is worth noting that these last two categories combined were chosen less frequently than any of the other three categories.

Table 4.10. Text difficulty ratings, from student reading logs.

Rating	N	%
Too Easy	467	47.3%
Kind of Easy	118	11.9%
Just Right	323	32.7%
Kind of Hard	58	5.9%
Too Hard	22	2.2%
Total	988	100.0%

While these numbers and percentages are interesting, they only tell us that students in the sample generally believed the books they chose to be easy or just right for them. Part of the problem with using this rating method on its own is that there is no way to calibrate it with students' beliefs about how easy or how difficult a text really should be. For example, the term "just right" could mean different things to different students; some may think that a "just right" book is slightly challenging, while others may view a "just right" book as one they can read easily and independently. In other words, one student's "just right" may be another student's "too easy," or even another student's "kind of hard." As a result, it is therefore necessary to compare results from this scale with other data sources, looking for signs of agreement or disagreement. The next analysis allows us to compare students' difficulty ratings with their actual performance on passage-specific oral reading accuracy and comprehension measures.

#### *Perceptions Versus Performance*

The first step in comparing students' perceptions of text difficulty with their actual performance was to look at this relationship at a somewhat macro level, measuring differences in student ratings of frustration-level and independent-level texts. As mentioned earlier, because reading log data was not available for all of the 35 matched cases in the subsample, this current analysis relied on interview responses for information about student perceptions of text difficulty, using a scale item parallel to the one on the reading log. Table 4.11 shows the distribution of students' text difficulty ratings for both the frustration-level texts and their independent-level counterparts, based on interview responses.

Table 4.11. Ratings of text difficulty by frustration-level and independent-level texts.

Rating	Frustration		Independent		Total	
	n	%	n	%	n	%
Very easy	4	11.4%	12	33.3%	16	22.5%
Kind of easy	7	20.0%	6	16.7%	13	18.3%
Just right	7	20.0%	17	47.2%	24	33.8%
Kind of difficult	14	40.0%	1	2.8%	15	21.1%
Very difficult	3	8.6%	0	0.0%	3	4.2%
Total	35		36 <sup>a</sup>		71	

<sup>a</sup> One student could not decide between “kind of easy” and “kind of difficult,” because she said that some parts were easier than others. Her ratings therefore show up in both rating categories, resulting in an n of 36.

The data in this table reveal several interesting things. First, it appears that students are relatively aware of the difficulty of their frustration-level texts, with 48.6% rating them either “kind of difficult” or “very difficult.” This finding is particularly striking in when we compare it to ratings of independent-level texts, for which only one student gave a rating of “kind of difficult,” and no students gave a rating of “very difficult.” However, the fact that another 11.4% of the subsample students rated their frustration-level texts “very easy” reminds us that their perceptions may not always be accurate. Second, a large number of students appear to equate “just right” with a relatively high level of oral reading performance, with 47.2% of students rating their independent-level texts “just right.”

While these findings are informative, this analysis is limited by the fact that it reduces reading performance to oral reading accuracy. Because it focuses on the

groupings of frustration- and independent-level texts, which were created based on oral reading accuracy alone, this analysis effectively equates reading performance with oral reading accuracy. To get a better picture of relationships between perceptions and performance, it is necessary to bring comprehension into the mix. To do this, the interview data are once again useful for providing more detailed information about students' perceptions of text difficulty. In addition to the scale item described above, the interview protocol included two items that addressed the separate components of word recognition and comprehension. These items asked students if there were any words they did not know or any parts of the text they did not understand; they responded by choosing from a 4-item scale: none, not very many, some, or a lot.

First, I looked at students' responses to the interview question, "Were there words that you didn't know?" By comparing students' responses for frustration- and independent-level texts, we can see that students' assessments of how well they knew the words in the text differed quite a bit across text difficulty levels (see Figure 4.3). For the frustration-level texts, more than half of the students (60.0%) in the subsample reported encountering either "a lot" (12.0%) or "some" (48.0%) words they did not know, and none of them reported finding "none." In contrast, for the independent-level texts read by the same subsample of students, 57.6% reported encountering "none," and another 36.4% said that there were "not very many" words they did not already know. Only 6.1% of the subsample students reported finding "some" unknown words in their independent-level texts, and none of them reported finding "a lot."

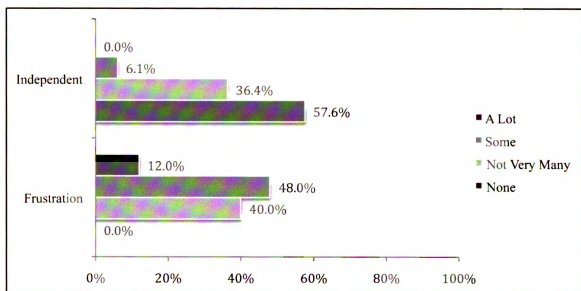


Figure 4.3. Responses to the interview question, “Were there words that you didn’t know?” by text difficulty.

Next, I conducted a parallel analysis of students’ responses to the comprehension-related interview question, “Were there parts of the book/magazine/other that you didn’t understand?” Subsample students’ responses were compared across the groupings of independent- and frustration-level texts, and the results are shown in Figure 4.4. The distribution of responses in this graph is both strikingly similar across the text difficulty groupings and strikingly different from the word-level data presented above. As this figure shows, despite the significant differences in actual oral reading accuracy and even in comprehension across the two sets of texts, students’ responses as a whole were not much different when it came to their perceptions of how well they understood the texts in question. In other words, it appears that the students in this sample did not always know what they did and did not understand from the texts they chose to read.

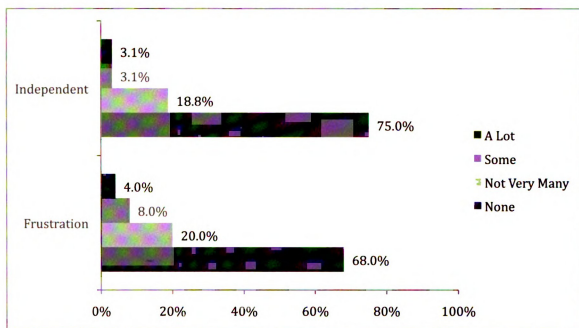


Figure 4.4. Responses to the interview question, “Were there parts of the book/magazine/other that you didn’t understand?,” by text difficulty.

These results are particularly interesting when they are considered in conjunction with the earlier finding that students’ perceptions of themselves as readers (from the MRP Self-Concept subscale) correlated significantly with their performance on the GMRT-4 Word Decoding subtest but not with their performance on the GMRT-4 Comprehension subtest. It would appear that the second grade readers in this sample were much more aware of their difficulties with word reading than they were of their difficulties with comprehension.

Another possibility is that a student’s familiarity with a text may influence his perceptions of that text’s difficulty; the role that prior knowledge plays in reading should not to be overlooked (e.g., Baldwin, Peleg-Bruckner, & McClintock, 1985; Kintsch & Franzke, 1995; Voss & Silfies, 1996; Wolfe & Mienko, 2007). To this end, the interview question related to previous experiences with a text offers some helpful insights. In particular, given the prominence of familiarity as a reason that students gave for choosing



frustration-level texts, it seems that a closer examination of the relationship between familiarity and perceptions of difficulty might be informative. This last analysis thus focuses on the subsample students' perceptions of difficulty and their prior experiences with the frustration-level texts that they chose to read.

Of the 35 students in the subsample, 18 reported that they were reading their frustration-level texts for the first time, while the remaining 17 said that they had read their texts anywhere from 1 to 100 times previously. Of the 18 students who were reading the texts for the first time, 3 rated their frustration-level text to be "very difficult," and 9 rated them "kind of difficult." Two students estimated the text to be "just right" for them, 2 rated them "kind of easy," and the remaining 2 rated their texts "very easy." Table 4.12 shows a distribution of students' difficulty ratings for their frustration-level texts, according to the number of prior experiences reading the text.

Table 4.12. Difficulty ratings of frustration-level texts, by number of prior readings.

Rating	0	1-2	5+
Very easy	2	2	0
Kind of easy	2	4	1
Just right	2	4	1
Kind of difficult	9	3	2
Very difficult	3	0	0

Students' interview responses also revealed their recognition that prior experience can make a text easier to understand. For example, the two students who were reading texts for the first time and who rated them "very easy" both commented that, although they had not previously read the books on their own, they were already familiar with the

story in some way. One of these two students was the boy, mentioned earlier, who attributed the perceived easiness of his chosen text (*Star Wars Episode I Journal: Anakin Skywalker* [Strasser, 1999]) to the fact that that he had seen the *Star Wars* movies before: “It was easy because I’ve saw all the movies, and I knew what was going to happen.” In the other instance, a girl who read *Bedtime Stories for Dogs* (Jasheway, 1996) rated it “very easy” and commented that the classroom aide had read the book aloud to the class before. Another student, who rated *Goldilocks and the Three Bears* (Bryant, 1995) “kind of easy,” mentioned that she already knew the story because she had once played the role of the baby bear in a play: “I want to read this book because I learned all about it in the play.” As these examples illustrate, students’ familiarity with specific, difficult texts – either through repeated reading or through other experiences with the same text or a similar text – may influence their perceptions of the text’s difficulty, independent of their actual reading performance.

#### *Question 4: Summary*

The focus of Question 4 was on students’ perceptions of text difficulty, particularly in relation to their actual performance. Students’ overall ratings of difficulty differed between the groups of frustration-level and independent-level texts, suggesting that students’ perceptions may be somewhat accurate, at least in relation to oral reading performance. However, an examination of more specific ratings related to word knowledge and comprehension showed markedly different results. Students seemed much more aware of their word recognition difficulties than they were of their comprehension difficulties. A closer look at students’ interview responses also suggests that familiarity with a text may mediate student perceptions of text difficulty.

### Question 5:

What are students' affective experiences with reading difficult texts?

Are frustration-level texts really frustrating, or do students enjoy them, particularly in comparison to the easier texts that they also choose to read?

Although one could certainly argue that whether or not a student enjoys a text is less important than whether or not he understands it, students' affective experiences with texts are also important. Enjoyment of particular texts and of reading in general contribute to motivation to read and to reading volume (e.g., Pressley et al., 2003; Pressley, 2006), which are both correlates of reading achievement (e.g., Guthrie et al., 1999; Stanovich, 1986). In this study, evidence of students' enjoyment comes from reading log entries and student interviews. These data sources offer information about students overall enjoyment of the texts they chose, and they also allow for insights into possible relationships between enjoyment and text difficulty. As with the previous analysis of perceptions of text difficulty, it is first helpful to look at the sample as a whole before moving to the subsample of matched cases.

#### *Overall ratings of enjoyment*

In general, students' ratings of their enjoyment for all of the texts they read were extremely high. In students' reading logs, which asked students to circle from one to five stars to indicate their enjoyment, ratings were overwhelmingly clustered near the high end of the scale, with 71.3% of chosen texts receiving the highest rating. Table 4.13 shows the distribution of these ratings for all reading log entries. In addition to the expected range of responses, which are included in the table, a small number of students

opted to expand the scale in both directions, with 10 ratings of 0, a rating of 6, and a rating of 10.

Table 4.13. Ratings of reading enjoyment, from student reading log entries.

Rating	n	%
1	93	9.6%
2	37	3.8%
3	53	5.5%
4	96	9.9%
5	692	71.3%
Total	971	100%

### *Enjoyment and Performance*

One of the purposes behind this fifth research question was to test Betts' assertion that difficult texts are frustrating for students to read. As the term "frustration level" implies, there is assumed to be a direct connection between students' cognitive performance on measures of oral reading accuracy and comprehension and their affective responses to their reading experience. To test this connection, I compared subsample students' ratings of enjoyment from the interview data between the groups of frustration-level and independent-level texts. I then took a closer look at the interview responses to search for more qualitative data that might illustrate the nature of the relationship between reading performance and reading enjoyment.

The first analysis was conducted at the group level and consisted of a comparison of students' responses to the interview question, "Did you enjoy reading this

book/magazine/other?” Students had been instructed to choose a response from a 3-item scale: not at all, a little, or a lot. Out of the subsample of 35 matched cases, for their frustration level texts, 6 students (17.1%) said that they enjoyed the text “a little,” and the remaining 29 students (82.9%) said that they enjoyed the text “a lot.” No students reported enjoying the frustration level text “not at all.” In comparison, out of the same subsample, 2 students (5.7%) said that they enjoyed their independent-level texts “not at all,” 4 students (11.4%) said that they enjoyed the text “a little,” and the remaining 29 students (82.9%) reported enjoying the text “a lot.” A summary of these results in presented in Table 4.14.

Table 4.14. Text enjoyment ratings for the subsample of frustration-level choosers, by text difficulty group.

Rating	Frustration		Independent		Total	
	N	%	N	%	N	%
Not at all	0	0.0%	2	5.7%	2	2.9%
A little	6	17.1%	4	11.4%	10	14.3%
A lot	29	82.9%	29	82.9%	58	82.9%

These results indicate that, as a group, subsample students’ enjoyment of their frustration-level texts was at least comparable, and even slightly higher, than their enjoyment of the matched independent-level texts, at least when based on oral reading accuracy. To see whether this finding held true for comprehension, I also sorted comprehension scores by enjoyment rating and looked at the ranges of comprehension scores that corresponded to each rating category. Table 4.15 shows the results of this analysis. Mean comprehension scores were comparable across the enjoyment categories

of “a little” and “a lot, while the two texts enjoyed “not at all” had relatively high comprehension.

Table 4.15. Text enjoyment ratings and comprehension scores for the 70 texts chosen by the subsample of frustration-level choosers.

Rating	n	Comprehension Scores		
		Min	Max	M
Not at all	2	75.0%	100.0%	87.5%
A little	10	41.7%	100.0%	70.7%
A lot	58	25.0%	100.0%	72.7%

In explaining their enjoyment ratings for their frustration-level texts, students gave a variety of reasons, including several that were directly related to the difficulty level of their chosen text. For example, one girl explained her statement that she liked *Carlita Ropes the Twister* (Canetti, 1999) “a little” by saying, “I liked that the girl saved her city. I didn't like that it had big, hard words.” In a similar vein, another girl said that she enjoyed *The Mind of the Cat* (Brotsky, 1990) “a little,” commenting, “it was kind of a little hard, and I liked it because it gave me so much information about it.” And a third girl, after reading *Dot the Fire Dog* (Desimini, 2001) with an oral reading accuracy rate of only 79.3%, reported liking the book “a lot” and agreed that she enjoyed the story even though the words were so hard. However, she also said that she would not choose another book like it again because, “It’s not the level I kind of read; it’s higher.” These comments imply that for some readers and some texts, their overall enjoyment of a text may be a combination of both their interest in a topic or a story and the difficulty of the text for them. Notably, five of the students who reported having enjoyed their frustration-level

texts “a lot” specifically mentioned that they liked the pictures. This finding suggests the possibility that students may genuinely be able to enjoy their interactions with a text despite relatively low levels of oral reading accuracy.

Regarding their independent-level texts, the students in the subsample also made a few comments related to text difficulty. Interestingly, for the two students who reported having enjoyed their chosen texts “not at all,” both of them gave explanations that were directly related to the easiness of the book for them. For example, one boy chose the decodable text *Hide and Seek* because “I thought it would be a really good book.” After reading it, however, he explained his “not at all” enjoyment rating by saying, “it was so easy, and I like harder books.” Another boy read *Clifford and the Big Leaf Pile* (Page, 2000) – which he had chosen from the selection of texts I provided to him – and reported enjoying it “not at all.” As he gave his detailed explanation for this rating, he pointed back and forth between the independent-level *Clifford* book and the frustration-level text he had chosen earlier that day: an instruction manual for the Wii version of the video game *The Legend of Zelda: Twilight Princess* (Nintendo, 2006):

Because it's for babies. Babies can read books like this [points to Clifford book], but babies don't know how to read like this [points to Zelda book] and this is for schools [Clifford], and this isn't [Zelda]. And this is rated T [Zelda], and that's rated E [Clifford].

His use of the standard video game rating system – T for Teen, E for Everyone – suggests his preference for texts aimed at older audiences and his dislike of books that are too easy, or at least that appear to be too easy. A third student read *Green Eggs and Ham*

(Seuss, 1960) and said that she liked it “a little” explaining, “I enjoyed about it that it's funny, and I did not enjoy about it that it's just too easy.”

These examples from the interview data present a mixed picture of the possible relationship between text difficulty and enjoyment. Based on the distribution of students' enjoyment ratings, there did not appear to be strong differences between their enjoyment of the difficult and easier texts they chose to read. However, a closer look at students' open-ended responses shows that a number of students did view text difficulty as directly influencing their enjoyment of a text in both directions (i.e., making the text either more or less enjoyable). To provide some quantitative data to supplement these findings, a Spearman rank-order correlation was conducted to determine the nature of the relationship between subsample students' perceptions of text difficulty and their enjoyment of the texts. This correlation method was used because both variables were ordinal, representing data from scale items on the interview protocol. The difficulty perception item (“How difficult or how easy was this book for you?”) and the enjoyment item (“How much did you enjoy this book?”) had a non-significant correlation of -0.044 ( $p = .718$ ,  $n = 69$ ), suggesting that for this subsample of students, enjoyment and perceived difficulty were not significantly linked.

#### *Question 5: Summary*

The focus of question five was on students' enjoyment of the texts they chose, particularly on ways that this enjoyment might be related to text difficulty. Overall, students' ratings of their chosen texts were very high, with the vast majority receiving the highest rating on a 5-star scale. A group-level comparison of rating frequencies for frustration-level and independent-level texts showed no clear relationship between



perceived text difficulty and text enjoyment. A review of interview responses showed that students do mention text difficulty as affecting their enjoyment of texts at both difficulty levels. Considering this finding in conjunction with the earlier results for Question 2, it appears that text difficulty can function as either an attraction or a deterrent, both for difficult and for easier texts. Finally, a Spearman correlation showed no significant relationship between students' perceptions of text difficulty and their ratings of text enjoyment. In other words, it appears that in nearly all cases the frustration-level texts are not actually frustrating for the students who chose to read them.

## CHAPTER 5: DISCUSSION

This chapter elaborates on the results presented in the previous chapter, making connections to relevant theory and previous research and offering some possible explanations for some of the main findings. The discussion ties together results from the five research questions into several key points, which serve as the foundation for a set of implications for classroom practice. The final sections of the chapter then address some important limitations of the study and offer suggestions for further research.

### Betts' Criteria: Too Stringent for Second Grade Readers?

As described in Chapter 2 of this dissertation, Betts' framework of reading levels has received substantial criticism over the years, most noticeably in relation to the development and validity of the levels' oral reading and comprehension criteria. The findings of this study lend additional support to the ideas of some of Betts' critics, particularly regarding the applicability of the criteria for younger readers and the combined requirements for oral reading accuracy and comprehension.

First, this study supports the contention of Powell (1970) and others (Johns & Magliari, 1989; Schummers, 1956) that the some of Betts' criteria may be too stringent, at least in the case of younger readers like the second graders in this sample. Even when neither self-corrections nor repetitions were counted as errors – a more generous scoring method than is used in nearly any commercial IRI – only 23 of the 159 assessed readings (14.5%) met the 99% independent-level criterion, even with rounding up any scores of 98.5% or higher (without rounding, only 14 of the 159 readings were at 99.0% or higher). Using a less generous (but more typical) scoring system of counting self-corrections as errors, only 7 of the 159 readings (4.4%) met the 99% criterion. These findings suggest

that the traditional independent-level threshold may be too high for second grade readers. It is worth noting that the original criteria were developed based on a single research study (Kilgallon, 1942) with a small sample (n=41) of fourth grade students, although they were intended by Betts (1946) to apply across the elementary grades. They were also established using a silent-before-oral method, which may have inflated comprehension scores (Kasdon, 1970; Powell, 1970). Given this information, it seems like Powell may have had the right idea in arguing for the creation of different criteria for different ages and grade levels.

In addition, Pehrsson (1994) has argued more generally for a reconsideration of Betts' leveling system, on the grounds that it is too conservative and that it does a disservice to readers by placing them with reading materials that are too easy. He comments:

Although frustration is a real possibility and true frustration should be avoided, the other side of the coin is that our assessment approaches may be overprotective. Is it possible that we have failed to challenge many students and thereby contributed to their slower progress? ... Is it possible that by exercising too much caution we may have caused our more mature students to dislike reading because they must read texts they consider to be below their intellectual potentials and far removed from their interests? ... Although it may be common sense to protect children from frustration, is it not also reasonable to challenge, to support, and to teach students to do something they cannot presently do? Is this not what education is about? (p. 203).

In this study, students were not frustrated by their frustration-level texts, and some of them also demonstrated good comprehension despite frustration-level degrees of oral reading accuracy. An investigation of the impact of text difficulty on reading achievement was beyond the scope of this study, but further research along these lines may provide some informative answers to Pehrsson's important questions.

Second, there was a considerable lack of correspondence between frustration-level and independent-level designations based on oral reading accuracy percentages and performance on comprehension questions. For example, of the 35 frustration-level texts, with oral reading accuracy percentages of 90% or less, only 14 had frustration-level comprehension rates of 50% or less. And of the matching independent-level texts, with oral reading scores of 92% or higher, 15 had independent-level comprehension rates of 90% or above. Although there was a strong relationship between oral reading accuracy and comprehension in general, this relationship not appear to hold for frustration-level texts, and there were a number of cases that don't quite fit the Betts model. This finding may be related to the first point, in that it could reflect a lack of fit between Betts' criteria and developmental reading behaviors. For example, Powell (1970) found that first and second grade students could tolerate an average of 85% oral reading accuracy while maintaining 70% comprehension, while third through sixth graders required a higher average threshold of 91-94% accuracy in order to maintain the same 70% comprehension level. As Powell (1970) explained,

The data suggest that the younger child can tolerate more word-recognition error and maintain an acceptable comprehension level than youngsters in grades three through six. Whether this difference is due to the complexity of the language used

for reading between these two groups, the difference in the depths of concepts presented in the reading materials at the upper levels, both language and concepts, or other factors not immediately discernible can only be verified through further research. (p. 107)

Findings from this current study appear to support the possibility that oral reading accuracy and comprehension for second grade students may not fit the pairs of criteria originated by Betts. It is possible that this lack of fit can be resolved by conducting further research to determine developmental gradients for paired criteria. However, it also seems possible that the lack of fit may be due not merely to developmental differences but to an uncertain relationship between oral reading accuracy and comprehension more generally. In general, given the continuing lack of clear evidence about the overall validity of Betts' criteria and their particular applicability for students of varying ages, serious work needs to be done in order to justify their continued use for placing students in reading level groups and for matching students with instructional and recreational texts.

#### Oral Reading and Comprehension: Uncertain Relationships

Additional analysis was intended to determine the extent of the connection between text-level comprehension and word-level difficulties. For the 70 assessed texts in the subsample, there was a statistically significant (.472) correlation between students' oral reading accuracy and comprehension, based on passage-specific comprehension questions. However, this correlation is actually quite low (indicating a 47.2% probability that relationships are due to something other than chance), especially for two aspects of reading that are frequently assumed to be closely related. One possible explanation for

the lack of a stronger correlation is that familiarity with a text may play a differential role in oral reading accuracy and comprehension. The methods used in this study necessarily meant that students had some prior knowledge of their chosen texts before participating in the individual assessments, because students were assessed on texts they had already read independently. One previous study on this topic found that having students read silently before reading orally resulted in significantly better comprehension after the oral reading, although it did not result in significantly better oral reading accuracy or faster oral reading rate (Kasdon, 1970). A similar effect may be present in this study: because students had already read their chosen texts silently at least once, their prior knowledge of the text may have aided their comprehension, although it would not necessarily have benefited their oral reading accuracy. For example, Kasdon (1970) found little relationship between decoding and comprehension in his study, offering the tentative explanation that “students seldom figure out pronunciations of words while reading silently unless the words interfere with comprehension” (p. 91). Based on this idea, it seems reasonable that students’ prior experiences with texts may have increased their chances of understanding texts without a concomitant improvement in their oral reading accuracy.

Regarding text difficulty and comprehension, it is also worth noting that only 14 of the 35 frustration-level readings (40.0%) had comprehension scores that fell into Betts’ frustration-level comprehension range of 50% or less. In fact, 6 of the frustration-level readings (17.1%) even met the 90% independent-level criterion for comprehension. The fact that a majority of the frustration-level readings had better than frustration-level comprehension, combined with the lack of consistent correlations between oral reading

accuracy and comprehension, suggests that poor oral reading accuracy does not necessarily mean that the student does not understand the text at hand. At the same time, good oral reading accuracy also does not necessarily imply good comprehension, as evidenced by the fact that 11 of the subsample students (31.4%) had comprehension scores of 75% or less while reading with at least 92% oral reading accuracy, and as high as 97.7% accuracy. This finding suggests the presence of some so-called “word callers” (e.g., Riddle Buly & Valencia, 2002), who, because of the developmental nature of the relationship between comprehension and oral reading accuracy, are able to pronounce words accurately without comprehending. As Paris and Hoffman (2004) explain:

It also appears that comprehension is more highly related to oral reading accuracy and rate in beginning readers and that the relation decreases by the time children are reading texts at a third- or fourth-grade level. This means that some children become adept “word callers” with little evidence of comprehension, so reading rate and accuracy measures in IRIs may yield incomplete information for older readers. (p. 207)

Given the prominent role that oral reading accuracy plays in determining reading levels, it seems that we may want to be more cautious about making assumptions regarding students’ comprehension based on oral reading performances.

Differences between oral reading and comprehension were also apparent in subsample students’ perceptions of text difficulty. The main finding on this point is that students were more aware of their word decoding difficulties than they were of their comprehension difficulties. This phenomenon was apparent in the lack of difference in students’ ratings of their comprehension difficulties for frustration-level and independent-

level texts, as compared to their ratings of their word-level difficulties for the same texts. It was also evident in the statistically significant correlation between GMRT-4 Word Decoding subtest scores and Motivation to Read Profile Self-Concept subscale scores, as compared to the non-significant correlation between the GMRT-4 Comprehension scores and the same MRP subscale. Based on those measures, students with poor word decoding performance were more likely to view themselves as poor readers; this relationship did not hold true for students with poor comprehension scores.

### The Frustration Level: A Misnomer?

One of the central assumptions in Betts' (1946) original framework was that the act of reading texts at high levels of difficulty would trigger feelings of frustration in young readers; hence, the term "frustration level." Recommendations against having students read difficult texts are often based on the similar assumption that frustration-level texts will lead to frustration and possibly to a decreased motivation to read and a subsequent avoidance of reading (e.g., Allington, 2006). However, one important finding from this study is that the level of difficulty of a text, at least in terms of word recognition, was not associated with a student's enjoyment of the text. A comparison of enjoyment ratings for the frustration-level and independent-level texts chosen by the subsample of frustration-level choosers showed that students' oral reading performance was not clearly related to their enjoyment of the text. Interestingly, this same finding held true for students' perceptions of text difficulty, which a Spearman rank-order correlation showed to be not significantly related to enjoyment for the students in the subsample.

One important implication of the findings in this area is that, in nearly all cases, the frustration-level texts were not actually frustrating for the students who chose to read



them. Assuming that self-reports of enjoying a text “a lot” can be viewed as an indication of a lack of frustration, the subsample students’ were actually slightly less pleased by texts they viewed as “too easy” than by texts they viewed as “too hard.” Contrary to what one might expect from Betts’ (1946) description of frustration-level texts, easy texts were reported as having been enjoyed “not at all” more often than difficult texts were.

Students’ repeatedly expressed dissatisfaction with some of the texts that they deemed “too easy,” describing them variously as “boring” or “for babies.” This may be in part because it is more socially acceptable to say that something was too easy than to say that it was too difficult. Overall, however, the assumption that difficult texts necessarily lead to some degree of emotional distress (e.g., Betts, 1946; Davis, 1975; Davis & Ekwall, 1976) is not supported by the results of this study.

One important question that is largely unresolved based on the findings of this study is whether students’ overall enjoyment of their difficult texts is an indication of embracing challenge or a symptom of false beliefs about abilities and performance. In other words, is the fact that students were not frustrated by their difficult texts – and did not always perceive them as being particularly difficult –unqualified good news, or is it something with which we should be concerned? Motivation and enjoyment are based on perceptions, which are necessarily subjective, so it is not surprising that students’ perceptions of their performance and their ability may have exerted a greater influence on their enjoyment than their actual performances did. Should teachers work to make students more aware of their actual performance, even if it might mean damaging their enjoyment of reading?

## Frustration-Level Text Choices: Frequent and Widespread

One important finding from this study is that students frequently chose to read frustration-level texts during their classroom-based independent reading. The distribution of the 159 assessed texts across Lexile levels showed that 42.2% of them were at approximately a 3<sup>rd</sup> grade level or higher, suggesting that text choices with a readability level above grade level were a relatively common occurrence (although whether “above reading level” equates to frustration-level depends on the reading skills of individual readers). This finding is consistent with the findings from a number of previous works that also found high readability levels relative to grade levels for student-selected texts (G. Anderson, 1985; Kragler, 2000; Mork, 2000). This study aimed to extend the research base by finding out whether the high readability texts that students chose were actually frustration-level texts for them in terms of oral reading accuracy and comprehension. Text-specific measures were therefore used to determine the actual difficulty of particular texts for particular students, at least according to commonly used assessment methods. The result of these procedures was that 50% of the sample ( $n = 70$ ) was assessed reading a frustration-level text at some point during the data collection period. It is worth noting that the subsample of 35 students was identified after assessing only 2.27 texts per student, on average. It certainly seems possible that a longer data collection period would have resulted in an even higher percentage of students reading frustration-level texts at some point. This finding also points toward the need for more research in the area of text difficulty. Given that half of the participating students chose frustration-level texts during classroom-based independent reading, it is important to learn more about the nature of students’ cognitive and affective experiences with those difficult texts and the impacts

that the difficulty level of students' text choices may have on their motivation to read and on their developing reading skills. This study has made some beginning steps in this direction, but more research is needed, as discussed at the end of this chapter.

A related finding is that the practice of choosing frustration-level texts was evident among students across the range of all three student variables examined in this study: reading ability, motivation to read, and gender. Regarding gender, a greater proportion of girls was identified as reading frustration-level texts, although statistical analysis found that the gender proportions in the subsample of frustration-level choosers were not significantly different from the gender proportions of the sample as a whole. This finding of no statistical significance complements Olson's earlier (1984) finding of no gender effect related to choices of texts with high readability levels. The apparent, though not statistically significant, gender differences suggest some support for the idea that "girls view reading as an activity that has personal significance for them, whereas boys attach somewhat less value to reading per se" (Good & Brophy, 1977, p. 361). This statement is consistent with the finding that girls had higher scores on Value of Reading subscale from the MRP. In relation to the expectancy x value model, the apparent tendency for girls to choose frustration-level texts may explained by the higher value they placed on reading, enabling them to tolerate a higher degree of difficulty than would a student who values it less.

In terms of motivational profiles, the subsample of students who were assessed reading frustration-level texts was also representative of the larger sample, both for self-concept and value of reading scores on the MRP. This finding is somewhat surprising, given that the expectancy x value model – and its corresponding empirical foundation –

seems to suggest that there should indeed be a connection between certain motivational traits and the difficulty of tasks that individuals choose to pursue (e.g., Atkinson, 1964, Brophy, 2004; Weiner, 1972; Wigfield & Eccles, 2000). One possible explanation for the lack of an observed relationship between the practice of choosing difficult texts and the motivational constructs of self-concept as a reader and value of reading is that a trait-based measure like the MRP is not an accurate predictor of individual behavior in specific situations. In other words, even if the MRP self-concept and value of reading items serve as reliable indicators of general expectations of reading success and the value attributed to reading tasks, the combination of these motivational components still might not be a useful predictor of student behavior in specific reading situations. Therefore, the lack of any motivational differences between students who did and did not choose to read frustration-level texts may simply mean that there were no differences, or it could also imply a lack of correspondence between dispositional motivation as measured by MRP scores and situational motivation as seen in students' choices, purposes, and responses. A second possibility is that there may not have been enough variance in MRP scores – which were generally quite high – to accurately predict the relatively widespread phenomenon of choosing difficult texts. A final possibility, which will be explored further in a later part of this discussion, is that students' choices of difficult texts were influenced by optimistic views of reading competence and by a lack of awareness of reading difficulties, particularly related to comprehension.

In terms of reading ability, the fact that the subsample of frustration-level choosers included some very skilled readers from each classroom indicates that the classroom libraries all contained at least some texts that were challenging for skilled

readers. It also shows that even skilled readers sometimes chose to read texts that they could not read with a high degree of oral reading accuracy. For instance, one girl with near perfect GMRT-4 scores read the chapter book *Ghosthunters and the Gruesome Invincible Lightning Ghost* with an 87.1% oral reading accuracy rate, stumbling over relatively difficult words like “harmlessly,” “reserved,” and “discreet.”

However, while the subsample contained students with a wide range of reading abilities, students with low GMRT-4 scores were disproportionately overrepresented; on both GMRT-4 subtests, students who chose frustration-level texts had significantly lower percentile scores than their classmates who did not choose frustration-level texts. This finding complements the findings of a number of earlier studies (Donovan et al., 2000; Fresch, 1995; Kragler, 2000; Olson, 1984), which found struggling readers more likely to choose texts above their level in terms of readability. It is not surprising that poor readers are more likely to be found reading difficult texts. Even in a well-stocked library that offers texts at a wide range of reading levels, the number of available texts that a skilled reader can read with ease is necessarily larger than the number of similar texts for less-skilled readers; sheer probability dictates that poor readers will have a better chance of choosing texts that are difficult for them (Donovan et al., 2000). For example, for some of the lowest readers, it was very difficult to find any texts that they could read at even the 92% criterion for oral reading accuracy. One student was assessed reading four frustration-level consecutive texts without achieving anything higher than an 85.6% oral reading accuracy level. Several other struggling readers had similar experiences, with two or even three frustration-level readings before finally finding an independent-level match.

One problem with the argument that poor readers are more likely to choose difficult texts because the library contains more texts above their level is that it depends largely on chance rather than on choice. In this study, students were free to choose from libraries that included texts of varying levels of difficulty, and they selected texts purposefully, not randomly. Fielding and Roller (1992) have suggested that students sometimes choose difficult texts because they do not have access to texts that are written at an appropriate reading level for them; this did not appear to be the case in this study. The classroom selection criteria were designed specifically to prevent against this possibility, and the classroom libraries all contained numerous lower level texts. The fact that struggling readers often chose to read difficult texts despite the availability of lower level texts suggests the need for an explanation other than lack of access. Some of the findings from this study indicate that students either were not aware of the difficulty of the texts they chose or were not interested in the easier texts available to them. An exploration of students' purposes for choosing particular texts offered important insights into possible reasons why so many students ended up reading frustration-level texts. Findings related to this part of the study are discussed in the following section.

#### Text Difficulty: A Secondary Consideration

Analysis of students' reading log entries and interview responses made it possible to test some of the hypotheses put forth by other researchers to explain why students choose texts beyond their tested reading ability. First, students often explained their choices of difficult texts with reasons related to interest in a particular topic, lending support to the hypothesis that interest in a topic and a desire for learning can lead students to take on challenging texts (Donovan et al., 2000; Hunt, 1970), which may be the only

avenues toward particular content (Brown, 2000). Second, students also mentioned choosing frustration-level texts because of prior experiences with the text or prior knowledge about the subject matter. For example, a number of the frustration-level texts in this study were books that teachers had read aloud, stories that students knew from having seen the movie version, or texts that provided additional information about current curricular topics. In these cases, text or topic knowledge may have motivated students to choose difficult texts by increasing their expectations of success. Indeed, familiarity appears to mediate students' perceptions of text difficulty, with subsample students generally less likely to rate texts as difficult if they were already familiar with the texts, whether through multiple independent readings, teacher read-alouds, or prior experiences with the content or story line. This finding supports Donovan and colleagues' (2000) hypothesis that familiarity may be an important factor in students' decisions to read difficult texts. Third, several students cited non-print features such as illustrations in relation to their frustration-level choices. Of the eight texts that students mentioned either choosing or enjoying because of the pictures they contained, six were frustration-level texts. This finding may support Moss and McDonald's (2004) idea that students choose to read difficult texts that contain non-print features that can facilitate understanding even in the presence of a large number of unfamiliar words.

Another important finding in this area was that difficulty level was rarely mentioned as a reason for choosing a particular text. This finding suggests that students' direct attention to text difficulty is secondary at best – it was mentioned infrequently and it rarely overruled a student's primary reason or reasons for choosing a particular text. For the few students who mentioned using text difficulty as a factor in making their text

choices, their efforts to either seek out or avoid a text based on difficulty were applied roughly equally to challenging and easy texts. For example, students mentioned avoiding easy texts about as often as they mentioned avoiding difficult texts. With few exceptions, students did not appear to choose frustration-level texts purely for the challenge of them. Rather, in pursuing their varied goals – such as finding information, reading what their friends were reading, exploring a classroom curricular topic, or reading about a favorite movie – students sometimes found themselves reading texts at what would typically be considered a frustration level in terms of oral reading accuracy. Findings in this area also offer some support to Fielding and Roller’s (1992) contention that some students do not want to read appropriate level texts. Books that were “too easy” were sometimes looked down upon, and one student even commented with disdain that an easy book he had just finished was “for babies.”

Another possibility is that students are simply not aware of the difficulty of their chosen texts, as measured in terms of word recognition. Students in the subsample of frustration-level choosers did show some awareness of text difficulty, with nearly half of the frustration-level texts rated either “kind of difficult” or “very difficult,” while only one of the independent-level texts received either of these ratings. However, this result also means that slightly more than half of the students rated their frustration-level texts as either “just right” or better. This finding has several possible explanations. First, as mentioned earlier, it is possible that the students were simply not aware of a given text’s difficulty for them. Second, it is possible that they recognized a certain degree of difficulty but felt that it fit within their personal definition of a “just right” text. Third, students may have intentionally given overly optimistic ratings as a way of preserving



their self-concept as a reader or as a form of social desirability bias, trying to present themselves in the best light to the researchers who would be reading their logs. Given the discrepancies between students' perceptions of text difficulty and their actual performance on text-specific oral reading and comprehension measures, it seems possible that students sometimes chose frustration-level texts because they were not aware of how difficult the texts actually were for them, at least in terms of word recognition. Fielding and Roller (1992) have suggested that students might not know how to find books at an appropriate level. The findings from this study indicate that students may have different notions of what constitutes an "appropriate level" and that they may not be aware of the difficulties they have with the texts, particularly in terms of comprehension. Taken together, these two ideas help contribute to our understanding of why students sometimes choose to read frustration-level texts during classroom-based independent reading time

#### Text Difficulty and Reading Performance: Mixed Results

One key finding in this area was that students who chose frustration-level texts generally demonstrated better comprehension of their independent-level texts than their frustration-level texts. This finding held true both for group averages and for case-by-case comparisons, in which 25 of 34 students (74.1%) had better comprehension of their independent-level texts. However, although students' comprehension of difficult texts was generally not as good as their comprehension of easier texts, this was not always the case. In fact, 2 of the students (5.7%) showed equal comprehension, and 7 students (20.0%) had better comprehension of their frustration-level texts. For the students whose comprehension of the frustration-level text exceeded that of the independent-level text, a closer look at interview data suggests that familiarity with the difficult texts may have

aided their comprehension. For example, three of the seven students mentioned that either the teacher or the classroom literacy aide had read their chosen text aloud to the class. Two others cited their own multiple, previous readings of their frustration-level texts: one student said that she had already read *Animal Poems* “lots and lots” of times, while the other claimed to have read the particular issue of *Nickelodeon Magazine* “at least a hundred times.” These previous experiences with the texts likely enabled the students to answer comprehension questions correctly despite their assessed difficulties reading the individual words in the selected passages. In addition, it is important to point out that even for the 25 cases in which comprehension was better for the independent-level text, comprehension for the frustration-level text was not always particularly low. In fact, in 11 of those cases comprehension scores were higher than 50%, which is the upper limit for the frustration level, and scores ranged as high as 91.7%. Although the very idea of quantifying comprehension is difficult and problematic, the findings from this study also suggest a need for learning more about how much comprehension is enough to make a reading experience worthwhile.

### Limitations

While the classroom setting ensured a diversity of texts and increased the chances that students’ text choices would reflect their actual preferences and purposes, the resulting text diversity posed some significant challenges. For example, the same assessments were applied to a wide range of student text choices, including graphic novels, poetry collections, picture books, informational texts, magazines, rhyming books, and chapter books. However, not all of these texts were equally well suited to the traditional assessments of comprehension and oral reading accuracy. Differences in text

type made it difficult to apply readability formulas, select running record passages, score oral retellings, and generate comprehension questions. The use of text-specific comprehension questions for each student necessarily introduced some variability into the results, although the use of a framework for generating and selecting questions was intended to aid reliability by promoting both consistency across texts and students and comparability with published reading inventories. In addition, text diversity also led to some apparent differences in oral retellings, possibly due to differences in the amount and connectedness of text content: a retelling for a book of 50 poems is qualitatively different than a retelling for a narrative picture book. Findings from another study suggest the possibility that the diversity of texts may have also played a role in students' comprehension. Moravcsik and Kintsch (1993) found that the way a text was written (in terms of good, well-organized writing versus poor, disorganized writing) had a significant influence on students' recall, independent of the influences of prior knowledge and reading ability. In other words, the way the text was written exerted a significant, independent influence on students' comprehension. These challenges suggest a need for more flexible assessment methods that can accommodate the diversity of texts that students read and understand, beyond the traditional methods that assume or require certain characteristics of a text's structure and content.

The second limitation relates to the degree to which students' text choices for this study were representative of their actual choices during the rest of the school year. First, it is possible and perhaps even likely that students' choices were affected to some degree by the knowledge that their text choices were going to be under close investigation by outside researchers. The text choice data may therefore reflect some social desirability

bias in this area, given the overwhelmingly positive response we regularly received upon entering the participating classrooms. Especially after the first several visits to a classroom, students greeted us warmly and were generally eager to participate in the one-on-one assessments. A number of students approached us to ask if they could read to us and to proudly show us the completed pages in their reading logs. Second, the very fact that teachers were asked to encourage their students to choose freely from the classroom library may have prompted students to choose texts that they would not otherwise have chosen for independent reading time, at least for classrooms in which free choice was not standard practice. Also, although teachers were asked to encourage students to choose freely, it is unclear how consistently they did so, since there was no observational data taken during independent reading time.

A third limitation of this study involves the reliability of the reading log data. The results of the reading log analyses may be somewhat skewed by the fact that a number of students gave similar ratings for all of their reading log entries. For example, one girl had 21 entries in her log, with all of them rated “just right.” Another girl had 43 entries, all of them rated “too easy.” These patterns of entries have two distinct implications for this study. First, the practice of giving identical ratings for every text suggests that students’ log entries may not be entirely reliable as sources of information. Self-report bias is an unavoidable issue when trying to assess motivational variables like goals and perceptions; it is therefore important to point out that the observed patterns in students’ ratings may represent actual differences in perceptions, or they may simply be an artifact of the way the students responded to the reading log prompt. Second, the distribution of responses suggests that students may operate under different concepts about how difficult

or how easy a text should be for them. For instance, the fact that a number of students appeared to use either “too easy” or “just right” as a default response suggests that the choices on the scale may not have all resonated with students; different wording for the rarely selected scale items may have resulted in a different distribution of student responses. This issue may merit further investigation, in order to find out more about how students and teachers define “just right” books in terms of text difficulty.

### Implications

One of the central goals in designing this study was to help answer some questions that have direct implications for classroom practice. For this reason, it was particularly gratifying to hear the five classroom teachers describe their reasons for choosing to participate in the study. Several of them mentioned having wondered about similar questions themselves: Should I give my students free choice for independent reading? Are they choosing things that they can actually read and understand? I know they enjoy that book, but will it help them become a better reader? What do I do about students who want to be chapter book readers like their friends, but aren’t yet ready for difficult books? The findings of this study do suggest some practical implications for classroom practice and assessment, which will hopefully prove useful to the five participating teachers and to others in similar positions.

First, the finding that students are less aware of their comprehension difficulties than they are of their word reading difficulties suggests that increased attention to comprehension monitoring may be in order. When guiding students to become self-regulated, independent readers, it may be helpful to increase the emphasis on comprehension and to teach students how to monitor their understanding of texts as they

read. In addition, teachers should consider going beyond the commonly used five-finger method when giving their students guidance on how to select texts. Although this method is easy to use, its effectiveness is currently untested, and it may lead students to focus solely on word recognition for determinations of appropriateness. A logical alternative would be to use a method that combines attention to word recognition, comprehension, and purposes for reading.

Second, because familiarity can aid students' comprehension of texts that are difficult in terms of word recognition, teachers should look for ways that they can scaffold students' understanding. For example, a teacher might consider reading difficult texts aloud before making them available in the classroom library. Another common teaching technique that receives some support from this study is the practice of providing thematically grouped texts linked to classroom curriculum. In this way, students' growing content knowledge in an area can help them understand more challenging texts than they might otherwise be able to manage on their own.

Third, findings related to students' purposes for reading suggest that we may need to think of success in independent reading more broadly than simply in terms of reading achievement. Students have varied goals for the texts they choose, all of which may be important to their enjoyment of the texts and to their long-term motivation to read. The RAND (2002) model of reading comprehension suggests that the consequences of reading activities include outcomes such as knowledge (gaining information), application (finding how to do something), or engagement (enjoyment). This model and the findings from this study remind us that students have individual goals for each text they choose, and their engagement with those texts can lead to a variety of outcomes. These

consequences may, in turn, influence students' future reading choices. As a result, text placement decisions should be made on a text-by-text basis rather than using only an ability-readability match. Success can therefore be redefined to include achievement of text- or activity-specific goals.

Fourth, the findings regarding Betts' reading level criteria suggest that teachers need to be critical consumers of IRIs and related assessments. In particular, it appears that the 99% independent level criterion for word recognition probably needs to be revised downward, at least for second grade readers. In addition, the practice of using IRI results to making reader-text matches should be broadened to take instructional goals and individual purposes into fuller consideration. If a teacher's goal for independent reading is for students to have enjoyable reading experiences, then readability matching should be de-emphasized, because "too easy" can be a larger barrier to enjoyment than "too difficult," at least for self-selected texts. If a teacher's goal is comprehension, then text difficulty may demand greater consideration, although teachers can also scaffold students' comprehension of difficult texts (in terms of word recognition) by building their familiarity with the text and the content. In general, the lack of adequate reliability and validity information for Betts' framework and related assessment methods calls out for additional research and requires us to be cautious when using them to assess students' reading performance and to make reading level placement decisions.

Fifth, the important finding that students were not frustrated by their frustration-level texts suggests that teachers should carefully consider the assumptions that underlie the frequent recommendations against having students read difficult texts. At least in the

context of independent reading, it appears that the term frustration-level may be a misnomer. As Pehrsson (1994) comments:

I am deeply concerned that we have accepted the concept of frustration level without question and thereby have possibly hindered many students by not sufficiently challenging them. I suggest that we challenge, support, and teach many students to read texts that traditionally would be labeled at their frustration level and avoided. To describe texts that students can read only after instruction, we should replace the term *frustration* with the term *challenge*. (p. 207, emphasis in original).

Teachers should also be aware of the possible motivational consequences of having students read easy texts, given that a number of students expressed dissatisfaction with texts that they deemed too easy.

### Future Research

Although this study has made some contributions to our understanding of relationships between reading comprehension, motivation to read, and text difficulty, there are still a number of questions that remain to be answered. Some of these questions can be addressed through further analyses of the existing dataset, while others can serve as the focus of additional studies.

Further analysis of the data from this study could help address the issue of measuring text difficulty as a process rather than as an outcome. By conducting a more in-depth analysis of the oral reading data for the 159 cases in this study, it may be possible to find connections between variables like comprehension, reading rate, prosody, and error types, and to draw some conclusions about the types of reading behaviors that



may indicate effort or difficulty when comprehension is held constant. In addition, an analysis of error types on the running records could lend some insight into ways that types or patterns of errors may have differential effects on comprehension, at least for second grade readers. More qualitative work could be done on the motivational side, including a closer examination of the interview data and the construction of some general profiles and illustrative cases.

Additional studies are needed to further test the validity of current IRI assessment procedures and Betts' reading level criteria. In particular, given the finding in this study that the 99% independent-level criterion may be too high for second grade readers, it seems important to find out more about possible developmental differences in reading behaviors and to work toward a set of criteria that reflect any such differences. It also seems possible that there could be situational factors besides developmental level that affect either oral reading accuracy or comprehension, or both. For example, given Kasdon's (1970) finding that previous silent reading aided comprehension but not oral reading accuracy, it may be worth finding out if a similar effect exists for listening to a text read aloud before reading orally. Additionally, given the proven role that interest plays in comprehension and the importance of autonomy in motivational theory (e.g., Deci & Ryan, 2000), future research could also explore differences in reading behaviors on chosen texts as compared to assigned texts. Finally, while this current study touched on social and situational factors that influenced students' reading behaviors, additional qualitative research could do more to explore and describe the role that the sociocultural context plays in school-based text choices.

Another interesting, related line of research would be to find out more about why some students do not choose to read frustration-level texts. This study focused on the set of students who chose to read frustration-level texts, but important things may also be learned from their counterparts who did not choose to read frustration-level texts, or at least who were not identified as having done so during the data collection period. All of the classroom libraries in this study contained enough advanced chapter books and informational texts that every student probably could have found something at his or her frustration-level. Further investigation of the reading behaviors of these students may serve as an informative and useful complement to the findings from this current study.

Another issue meriting further research involves text difficulty as a general construct. As it is currently conceptualized and measured, the construct of text difficulty is somewhat problematic, particularly because it is an outcome measure rather than a process measure. The term “difficulty” usually implies effort, but in the case of reading it actually refers to proficiency. In other words, it is used to represent an individual’s actual reading performance on a particular text rather than the actual or perceived amount of effort that went into the reading of the text. The problem with this usage is that the very definition of text difficulty (as it is commonly used) requires limited success, while a successful reading performance necessarily indicates a lack of difficulty. What is missing in this conceptualization is a means of measuring effort and of distinguishing effortful reading from more fluent reading. In summary, the term “difficulty” appears to refer to the amount of effort that goes into the attempt, but the way it is used in relation to reading and text difficulty actually refers to the eventual success or failure of the outcomes.

Future research could work toward a more accurate measure of difficulty as related to effort.

On a related note, another assessment-related issue worthy of additional research involves the practice of generalizing student performance on samples of independent reading to instructional situations. The utility of IRIs for making reading level placements depends largely on the assumption that assessment of a student's independent reading can serve as an accurate predictor of the student's performance in both instructional and independent reading situations. For example, a certain level of performance reading orally and answering comprehension questions is said to indicate the level at which a student can read a text with instructional support. In other words, it seems curious that the methods typically used to determine the levels at which students can read with assistance or cannot read even with assistance involve reading without assistance. For instance, the very definition of the frustration level assumes that a certain amount of difficulty cannot be overcome even with assistance, although this assumption is never tested and the notion of assistance or instructional support is not elaborated. Further research could be used to examine the types of instructional support that might render more difficult texts readable, and it could also test the assumption that placement scores on IRIs can be generalized to performance on classroom texts in the presence or absence of instructional support.

And finally, perhaps the most pressing need for future research in the area of text difficulty is longitudinal. It is important to find out more about the impacts over time that reading texts at various levels of difficulty may have on a range of possible outcomes including reading comprehension, reading fluency, vocabulary knowledge, content knowledge, and motivation. Over time, what are the effects of spending time with texts

that can be read with various degrees of success in terms of word recognition and comprehension?

Answers to these questions and others can contribute greatly to our understanding of the important issue of reader-text interactions, in terms of relationships between text characteristics, student reading behaviors, and outcomes related to comprehension and motivation. Research in these areas should also lead to necessary improvements in methods of assessing reading ability and methods of applying assessment results to classroom practice.

## APPENDICES

Appendix A: Student Reading Log

Books I chose		Books I read	
<p>Title: _____</p> <p>_____</p> <p>_____</p> <p>_____</p>	<p>I chose this book because...</p> <p>_____</p> <p>_____</p> <p>_____</p>	<p>This book was...</p> <p><input type="checkbox"/> Too easy</p> <p><input type="checkbox"/> Kind of easy</p> <p><input type="checkbox"/> Just right</p> <p><input type="checkbox"/> Kind of hard</p> <p><input type="checkbox"/> Too hard</p>	<p>I give this book...</p> <p>☆ ☆ ☆ ☆ ☆</p> <p>because... _____</p> <p>_____</p>
<p>Title: _____</p> <p>_____</p> <p>_____</p> <p>_____</p>	<p>I chose this book because...</p> <p>_____</p> <p>_____</p> <p>_____</p>	<p>This book was...</p> <p><input type="checkbox"/> Too easy</p> <p><input type="checkbox"/> Kind of easy</p> <p><input type="checkbox"/> Just right</p> <p><input type="checkbox"/> Kind of hard</p> <p><input type="checkbox"/> Too hard</p>	<p>I give this book...</p> <p>☆ ☆ ☆ ☆ ☆</p> <p>because... _____</p> <p>_____</p>
<p>Title: _____</p> <p>_____</p> <p>_____</p> <p>_____</p>	<p>I chose this book because...</p> <p>_____</p> <p>_____</p> <p>_____</p>	<p>This book was...</p> <p><input type="checkbox"/> Too easy</p> <p><input type="checkbox"/> Kind of easy</p> <p><input type="checkbox"/> Just right</p> <p><input type="checkbox"/> Kind of hard</p> <p><input type="checkbox"/> Too hard</p>	<p>I give this book...</p> <p>☆ ☆ ☆ ☆ ☆</p> <p>because... _____</p> <p>_____</p>

## Appendix B: Comprehension Question Framework and Guidelines

### Question shells: narrative

What is this part of the story about?	Main idea	Literal
Who... What... (characters, objects)	Detail	Literal or Inferential
What was the first thing that happened? What happened after... What was the last thing that happened?	Sequence	Literal
Why did...	Cause and effect	Inferential
What is... What does        mean?	Vocabulary	Inferential or knowledge

### Question shells: expository

What is this section of the book about?	Main idea	Literal
What... How many... Give an example...	Details, facts, reasons	Literal
Why does... What happens when...	Cause and effect	Literal or Inferential
What is... What does        mean?	Vocabulary	Inferential or knowledge

### Criteria for good comprehension questions

(adapted from Valmont, 1972)

- Questions are asked in the same order as the information is presented in the passage
- Questions are as important as possible (central to passage meaning)
- Later questions are not answered by earlier ones
- Two questions do not call for the same answer (no overlap)
- Questions require specific answers (are not so broad that any answer is ok)
- Questions are as passage-dependent as possible
- Questions are answerable from the passage (not dependent on outside knowledge)
- Questions are not answerable from accompanying pictures
- Questions are specific, avoiding multiple, correct responses
- Questions are not leading, such that the question itself gives away the information that is being called for
- Questions do not offer 50/50 chances of success (e.g., yes/no)

Valmont, W. J. (1972). Creating questions for informal reading inventories. *The Reading Teacher*, 25(6), 509-512.

Appendix C: Sample Comprehension Questions

ID: 204	Title: Fizzkid the Inventor
---------	-----------------------------

What was that? Fizzkid looked up. She saw a big dinosaur. It was a pterodactyl. Fizzkid hid behind a tree. The pterodactyl didn't see her. But the pterodactyl saw the pogo stick. It grabbed the pogo stick and flew away.

"Oh no!" said Fizzkid. "I must get my pogo stick back."

The pterodactyl flew up to its nest. The nest was at the top of a very tall tree. Fizzkid climbed up the tree. She was scared. She didn't want the pterodactyl to see her, but she had to get her pogo stick back.

Fizzkid waited until the pterodactyl flew away. Then she peered into the nest. She saw three eggs.

Fizzkid stretched out her hand and grabbed the pogo stick. The pterodactyl was coming back. Fizzkid had to get away fast. She climbed down the tree as fast as she could. Fizzkid reached the bottom branch. She jumped into the sandpit below.

	Question	Answer	Score
1	When the pterodactyl came, where did Fizzkid hide?	Behind a tree	
2	What did the pterodactyl do when it saw the pogo stick?	It took the pogo stick and flew away up into the nest.	
3	How did Fizzkid feel when she climbed up the tree?	Um, I think it was scared.	
4	What does Fizzkid see when she peers into the nest?	Eggs	
5	Why did Fizzkid climb down the tree as fast as she could?	So she could get away from the pterodactyl	
6	What happened at the end of this section you just read?	She fell in the sand.	
TOTAL			
			%



## Appendix D: Student Interview Protocol

Name \_\_\_\_\_ Date \_\_\_\_\_

Title \_\_\_\_\_ Author \_\_\_\_\_

- 1) Why did you choose this book/magazine/other?
- 2) Had you read this book/magazine/other before? If so, how many times?
- 3) How easy or how difficult was this book/magazine/other for you?  
(scale: very easy, kind of easy, just right, kind of difficult, very difficult)
- 4a) Were there words that you didn't know?  
(scale: a lot, some, not very many, none)
- 4b) When you came to a word you didn't know, how often could you figure it out?  
(scale: always, almost always, sometimes, almost never, never)
- 5a) Were there parts of the book/magazine/other that you didn't understand?  
(scale: a lot, some, not very many, none)
- 5b) When you came to a part you didn't understand, how often could you figure it out?  
(scale: always, almost always, sometimes, almost never, never)
- 6a) The next time you choose something to read, how likely are you to choose another book/magazine/other like this?  
(scale: not at all, maybe, definitely)
- 6b) In what ways might you choose something similar?  
  
If no response, probe: For example, would you choose something by the same author or about the same topic?
- 7a) Did you enjoy reading this book/magazine/other?  
(scale: not at all, a little, a lot)
- 7b) If "a lot," What did you enjoy about this book/magazine/other?  
  
If "not at all," Why didn't you enjoy reading it?

## Appendix E: List of Provided Texts

- Brown, R., & Carey, S. (1994). *Hide and seek*. Illustrated by Sal Murdocca. New York: Scholastic.
- Desimini, L. (2001). *Dot the fire dog*. New York: Scholastic.
- Seuss, Dr. (1968). *The foot book*. New York: Random House.
- Seuss, Dr. (1960). *Green eggs and ham*. New York: Random House.
- Seuss, Dr. (1963). *Hop on pop*. New York: Random House.
- Lobel, A. (1979). *Days with frog and toad*. New York: HarperCollins Publishers.
- Lobel, A. (1976). *Frog and toad together*. New York: Harper & Row.
- Miles, E. (2003). *Wings, fins, and flippers*. Chicago: Heinemann Library.
- Nicholson, S. (1998). *A day at Greenhill Farm*. London: DK Publishing.
- Page, J. (2000). *Clifford and the big leaf pile*. New York: Scholastic.
- Smith, T. (2005). *Jake and the big fish*. South Melbourne, Australia: Cengage Learning.
- Wallace, K. (2000). *Wild baby animals*. London: DK Publishing.

## Appendix F: Oral Retelling Rubric

Oral Retelling Rubric

	0	1	2	Score
Sequence	Information related in a haphazard manner	Retains the general sequence of events; some elements recalled out of sequence	Retains sequence of events or steps	
Coherence	Bits of information related, but with little apparent organization	Relates an overall sense of the content; organization of the retelling is incomplete	Retelling is clear and coherent	
Accuracy	Contains inaccuracies	Contain minor misinterpretations or inaccuracies	Information is accurate	
Main idea	Leaves out key information	Retells most central or key events	Includes central or key events or points	
Detail	Details unrelated to larger points or ideas	Recalls some important facts; overlooks important details	Relates important facts and details	
			Total	/ 10

Adapted from:

Johns, J. L. (1997). *Basic reading inventory: Pre-primer through grade twelve and early literacy assessments* (7<sup>th</sup> ed.). Dubuque, IA: Kendal/Hunt Publishing Company.

# Appendix G: Running Record Score Sheet

Student ID: 0109	Text Title: Mister Seahorse
------------------	-----------------------------

## Accuracy

Mispronunciation	
Substitution	
Tester provided	
Insertion	
Omission	
Repetition	
Reversal	
Self-correction	
<b>TOTAL</b>	

## Rate

# of words	164
Time	
Rate	

## Meaning

Errors interfered with meaning:	
<input type="checkbox"/> Not at all	
<input type="checkbox"/> Slightly	
<input type="checkbox"/> Sometimes	
<input type="checkbox"/> Often	

## Fluency

1	2	3	4
<ul style="list-style-type: none"> <li>• Labored, disfluent reading</li> <li>• Very slow pace</li> <li>• All word by word</li> <li>• No phrasing</li> <li>• Awkward word groupings</li> <li>• No use of pronunciation</li> <li>• No intonation; monotone</li> </ul>	<ul style="list-style-type: none"> <li>• Choppy, plodding reading</li> <li>• Slow pace</li> <li>• Mostly word by word</li> <li>• Some 2 or 3 word phrasing</li> <li>• Appropriately grouped</li> <li>• Occasionally uses punctuation</li> <li>• Little intonation</li> </ul>	<ul style="list-style-type: none"> <li>• Fairly fluent</li> <li>• Reasonable pace</li> <li>• Some word by word</li> <li>• Mostly phrasing</li> <li>• Mostly meaningful phrases</li> <li>• Mostly uses punctuation</li> <li>• Some intonation</li> </ul>	<ul style="list-style-type: none"> <li>• Fluent reading</li> <li>• Good pace</li> <li>• Few word by word slow downs</li> <li>• Larger, meaningful phrases</li> <li>• Expressive interpretation guided by punctuation and meaning</li> <li>• Good, consistent intonation</li> </ul>

<p>As Mr. Seahorse drifted gently through the sea, he passed right by... A lionfish hidden behind a coral reef.</p> <p>But before long, Mr. Seahorse met another fish. "How are you, Mr. Tilapia?" asked Mr. Seahorse. Mr. Tilapia couldn't answer. His mouth was full of eggs. "I know, I know," said Mr. Seahorse. "Mrs. Tilapia laid her eggs. Now you are taking good care of them until they hatch." Mr. Tilapia nodded his head.</p> <p>"You must be very happy," said Mr. Seahorse and swam on his way.</p> <p>As Mr. Seahorse drifted gently through the sea, he passed right by... Several leaf fish hidden among the seaweed.</p> <p>But before long, Mr. Seahorse met another fish. "How are you, Mr. Kurtus?" asked Mr. Seahorse. "Perfectly fine," replied Mr. Kurtus. "Mrs. Kurtus laid her eggs and I have stuck them on my head. Now I am taking good care of them until they hatch."</p> <p>"You are doing a good job," said Mr. Seahorse and swam on his way.</p>
---

Appendix H: Running Record Error Scoring Matrix and Notes

	Substitution	Reversal	Omission	Insertion	Prompt	Self-correction	Proper Nouns	Repetition	Ignored Punctuation	Other
Analytical Reading Inventory (Woods & Moe, 2007)	Y	Y	Y	Y	Y	N	Y	Y	N	Hesitation: N
Basic Reading Inventory (Johns, 1997)	Y		Y	Y	Y	Y		N	Y	Hesitation: N Dialect: N
Critical Reading Inventory (Applegate, Quinn, & Applegate, 2008)	Y	Y	Y	Y	Y	N	N	N	N	Sub-vocalization: N Hesitation: N Dialect: N
Informal Reading Inventory (Burns & Roe, 1999)	Y	Y	Y	Y	Y	N	N	Y	N	
Reading Inventory for the Classroom (Flynt & Cooter, 2001)	Y		Y	Y	Y	N		N		
Reading Miscue Inventory (Goodman, Watson, & Burke, 2005)	Y	Y	Y	Y		Y		N	Y	Pause: N Dialect: N
Stieglitz Informal Reading Inventory (Stieglitz, 2002)	Y	Y	Y	Y	Y	N		Y		Dialect: N
Clay (1985)	Y	Y	Y	Y	Y	N	Y	N		Dialect: N
Kilgallon (1942)	Y	Y	Y	Y	Y			Y		Faulty Syllabication: Y
Lipson & Wixson (2003)	Y	Y	Y	Y	Y	N		Y	N	Hesitation: N

Analytical Reading Inventory (Woods & Moe, 2007):

- A repeated error on the same word is only counted as a single error if it doesn't change the meaning (e.g., "Mom" for "Mommy").
- A repeated error on the same word is counted in each instance if it changes the meaning (e.g., "tricks" for "trash").
- A repeated error on a proper noun is counted once if it is repeated consistently; it is counted multiple times if the substitution varies (e.g., "Dutch" and then "Bunch" for "Butch").
- Repeated non-word substitutions are always counted as errors each time they occur.
- Repetitions of parts of words are not counted as errors.
- Multiple repetitions of the same word or phrase only count as a single error, as long as the pronunciation doesn't change.

Basic Reading Inventory (Johns, 1997):

- A repeated error on the same word is only counted once (e.g., Bob for Bill)
- This same guideline also applies to situations where a nonsense name is used for a proper name or for any other word.
- Omission of an entire line counts as a single error.
- Only significant errors are counted. This number is determined by counting the total of all dialect errors, corrected errors, and errors that don't change meaning and then subtracting them from the total number of errors.

Critical Reading Inventory (Applegate, Quinn, & Applegate, 2008):

- The substitutions category also includes reversals.
- Omission of an entire line counts as a single error.
- A repeated error on the same word is only counted as a single error.

Informal Reading Inventory (Burns & Roe, 1999):

- Mispronunciations on proper nouns are not counted unless the names are also common words, like *Brown* and *Pat*.
- Omission of any continuous sequence of words counts as a single error.
- Repetition of any continuous sequence words counts as a single error.

Reading Miscue Inventory (Goodman, Watson, & Burke, 2005):

- Self-corrections of partial pronunciations are not counted.
- Multiple errors during a single correction attempt are only counted as a single error.
- Partial pronunciations that are not corrected are coded as omissions.
- Identical substitution or omission for the same printed word is counted as a single error.
- Varied responses for the same printed word are counted as separate errors.
- Substitutions and omissions of function words (e.g., determiners, verb markers, conjunctions, and prepositions) count each time they occur.

**Stieglitz Informal Reading Inventory (Stieglitz, 2002):**

- Nonword substitutions count as a full error, as do real (whole) word substitutions that disrupt the meaning of the text
- Real (whole) word substitutions that do not disrupt meaning count as a 1/2 point
- Omissions of one or more words count for a full point if they disrupt meaning; they count for 1/2 point if they do not disrupt meaning
- Insertions of one or more words count for a full point if they disrupt meaning; they count for 1/2 point if they do not disrupt meaning
- Repetitions of three or more words count as a full error; repetitions of two words or less do not count as a error

**Clay (1985):**

- For insertions of multiple, consecutive words, each inserted word is counted as a separate error.
- For omissions of multiple, consecutive words (including whole lines), each omitted word is counted as a separate error.
- Repeated errors count as errors on every occasion, except for proper nouns – for proper nouns, substitution of another proper name only counts the first time.
- Multiple errors in a phrase count as multiple errors, which means that a reversal of two words would count as two separate errors.

## Appendix I: Running Record Scoring Manual

### Mispronunciation

- **Description:** The reader attempts to pronounce a word but does so incorrectly, resulting in a nonword, and the error is not corrected by the reader.
- **Example:** The reader says “impersoned” for the written word “imprisoned.”
- **Scoring:** Each mispronounced word counts as a single error.

### Substitution

- **Description:** The reader reads a word incorrectly, resulting in an actual word, and the error is not corrected by the reader.
- **Example:** The reader says “thorn” for the written word “throne.”
- **Scoring:** Each substituted word counts as a single error.

### *Notes and special cases:*

1. **Proper names:** Any phonetically reasonable pronunciation of a proper name should be accepted as correct. For example, saying “Geronimo” with a long first ‘o’ is not counted as an error. In contrast, saying “Bunch” for the written word “Butch” should be counted as a substitution.
2. **Repeated errors:** If a student reads the same word incorrectly more than once in a passage, it only counts as one error, as long as the supplied word or nonword is the same in each case. For example, if a student says “theme” for the written word “them” in three places in a single passage, it is only counted as one error. However, if a student misread “Butch” once as “Dutch” and another time as “Bunch,” it should be counted as two separate errors.
3. **Partial pronunciations:** In general, only whole word substitutions and mispronunciations are counted. If a reader begins by only pronouncing the first part of the word, it should be coded depending on what happens next – the tester provides it, the reader self-corrects an incorrect first start, the reader correctly completes a correct first start, etc.

### Tester Provided

- **Description:** The assessor provides the word for the reader after the reader pauses for a long time or asks for assistance. This category only includes instances in which the assessor provides the correct word *and* the reader does not attempt it on their own.
- **Example:** The student sees the word “saucer,” pauses without attempting a pronunciation, and the assessor says, “saucer.” If the student had seen the word “saucer,” said “sugar,” and then the assessor says “saucer,” it should be coded as a substitution rather than tester provided.
- **Scoring:** Each tester provided word counts as a single error.



### **Omission**

- **Description:** The reader omits a word or words from the written text.
- **Example:** The reader says “the three pigs” for the written words “the three little pigs.”
- **Scoring:** Omissions of one or more consecutive words are counted as a single error, even for omissions of entire lines.

### **Insertion**

- **Description:** The reader adds a word or words that are not in the written text.
- **Example:** The reader says “all of the dogs” for the written words “all the dogs.”
- **Scoring:** Insertions of one or two consecutive words are counted as single errors. For insertions of more than two words, every inserted word beyond the first two is counted as a single error.

### **Reversal**

- **Description:** The reader reverses the order of two words in the written text.
- **Example:** The reader says “are you” for the written words “you are.”
- **Scoring:** Each two-word reversal counts as a single error.

### **Repetition**

- **Description:** The reader repeats one or more consecutive, whole words without changing or correcting any of them.
- **Example:** The reader says “the old woman, old woman, closed the door” for the written text “the old woman closed the door.” If the reader says “the woman, the old woman, closed the door,” it should be coded as a self-correction (of an omission) and not as a repetition. Similarly, if the reader says, “the old” and then pauses, the assessor provides the word “woman,” and the reader says “the old woman,” it should be coded as tester provided and not as a repetition.
- **Scoring:** Although repetitions are not considered errors and are not included in the error total, they should still be coded and tallied. Repetitions are not counted as errors. In fact, they are only coded as repetitions when they occur in isolation. They are not coded when they are part of a self-correction or when they are related to a word that has been provided by the tester.

### **Self-Correction**

- **Description:** The reader spontaneously corrects an earlier error, including an incorrect partial attempt.
- **Example:** The reader says “it’s time to knew, you knew” for the written words “it’s time you knew.” An example of a self-correction of a partial attempt would be if the reader says “hop, hoping” for the written word “hoping.” In contrast, saying “hop, hopping” for the written word “hopping” would not be coded as an error at all, since the partial attempt was correct and the reader then completed the word correctly.
- **Scoring:** Each self-correction of an earlier error counts as a single error.

## Appendix J: Leveling Criteria Matrix

Various sets of word recognition and reading comprehension criteria for independent, instructional, and frustration reading levels.

WR = word recognition; RC = reading comprehension

	Independent	Instructional	Frustration
Betts (1946) Woods and Moe (2007) Stieglitz (2002) <sup>a</sup>	99-100% WR 90-100% RC	95-99% WR 75-89% RC	90% or less WR 50% or less RC
Allington (2006)	98-100% WR 90-100% RC	95-97% WR 75% RC	95% or less WR 75% or less RC
Barr et al. (1995) <sup>a</sup>	98-100% WR 90-100% RC	95-97% WR 75-89% RC	90% or less WR 50% or less RC
Lipson & Wixson (2003)	96-99% WR 75-90% RC	92-95% WR 60-75% RC	90-92% or less WR 60-75% or less RC
Powell (1970) <sup>b</sup> oral rereading	94-100% WR 80-100% RC	92-94% WR 70-80% RC	91% or less WR 70% or less RC
Powell (1970) reading at sight	94-100% WR 80-100% RC	88-94% WR 55-80% RC	86% or less WR 55% or less RC
Leslie & Caldwell (2001)	98-100% WR 90-100% RC	90-97% WR (all Es) <sup>c</sup> 95-97% WR (MC) 70-89% RC	< 90% WR (all Es) < 95% WR (MC) < 70% RC
Johns (1997)	99-100% WR 90-100% RC	95-99% WR 70-90% RC	90% or less WR 50% or less RC
Applegate, Quinn, Applegate (2008)	99% WR 90% RC	95% WR 75% RC	90% WR 50% RC
Burns and Roe (1999)	99-100% WR 90-100% RC	85-99% WR (gr 1-2) 95-99% WR (gr 3-12) 75-90% RC	<85% WR (gr 1-2) <90% WR (gr 3-12) <50% RC
Caldwell (2002)	98-100% WR 90-100% RC	90-97% WR (all Es) 95-97% WR (MC) 70-89% RC	<90% WR <70% RC
Clay (1985)	Easy: 95-100% WR	90-94% WR	Hard: 80-89% WR
Armbruster, Lehr, & Osborn (2003)	95% WR	90% WR	<90% WR

<sup>a</sup> Barr et al. (1995) mention a category called 'borderline', which is defined by 90-94% word recognition and 50-74% reading comprehension. Similarly, Stieglitz (2002) includes a category called 'questionable', which is defined by 90-95% word recognition and 50-75% reading comprehension.

<sup>b</sup> Powell (1970) offers different criteria according to different grade levels (grades 1-2, grades 3-5, and grades 6+). The numbers in this table are for grades 1-2.

<sup>c</sup> Leslie and Caldwell use different thresholds for word recognition, depending on whether all errors are counted (all Es, above) or only errors that involve meaning change (MC, above).

## Appendix K: Coding Manual for Reasons for Text Choices

This coding manual was used to code student responses to two related items:

- Reading Log prompt: I chose this book because...
- Interview question 1: Why did you choose this book/magazine/other?

Coding Category	Description	Example(s)
Adult recommendation	The text was recommended by an adult (e.g., parent, teacher, librarian)	<ul style="list-style-type: none"> <li>• [My teacher] thought that the story would be interesting for me.</li> <li>• [My teacher] gave it to me</li> </ul>
Author	The text was by a favorite author	<ul style="list-style-type: none"> <li>• because I like Patricia Polacco's books</li> <li>• I like books written by Robert Munsch</li> </ul>
Character	The text was about a favorite character	<ul style="list-style-type: none"> <li>• I like Jack and Annie</li> <li>• I really liked the gingerbread man</li> <li>• Because I like Clifford</li> </ul>
Classroom connection	The text was related to a topic or genre that had been the focus of classroom instruction and learning	<ul style="list-style-type: none"> <li>• We are learning about tadpoles</li> <li>• It's kind of interesting to me because we're learning about things that swim in water</li> <li>• We are learning about dirt</li> </ul>
Difficulty	The text was written at a preferred level of difficulty	<ul style="list-style-type: none"> <li>• It's a little easier and not that hard</li> <li>• It looks easy</li> <li>• Some words are very, very hard for me to read and pronounce</li> </ul>
Entertainment	The text was perceived to be entertaining	<ul style="list-style-type: none"> <li>• Because it looked fun</li> <li>• Because I thought it would be funny</li> <li>• I just wanted to entertain myself</li> </ul>
Familiarity	The text was familiar, either through previous readings or experiences or through related content knowledge	<ul style="list-style-type: none"> <li>• Because I seen like cartoons of it on loony tunes, then I thought of reading it.</li> <li>• I have all the movies of it, so I just knew what happens and I haven't read this book before.</li> <li>• Ms. S. the aide read it to us.</li> </ul>

Format	The text was written in a preferred format	<ul style="list-style-type: none"> <li>• Because it has chapters</li> <li>• Because it got three stories in it</li> <li>• It is a play</li> </ul>
General	Vague, general comments	<ul style="list-style-type: none"> <li>• Because I thought it would be a really good book</li> <li>• It's a pretty good book</li> <li>• It's a nice book</li> <li>• It's cool</li> </ul>
Genre	The text was an example of a preferred genre	<ul style="list-style-type: none"> <li>• Because I like to read about fiction things</li> <li>• I like fairy tales</li> <li>• I like fiction books</li> </ul>
Improving reading	The text was perceived as helping improve reading skills	<ul style="list-style-type: none"> <li>• Sometimes it's a challenge for me to try to read things because I want to be a better reader. So in kindergarten I had little books and I got higher and higher, and I'm hoping next year that I might be able to read higher books</li> </ul>
Indecision	The text was chosen by default, because the student was unsure what to choose	<ul style="list-style-type: none"> <li>• I didn't know what book to pick</li> <li>• It been hard to pick</li> </ul>
Information	The text was thought to contain specific, desired information	<ul style="list-style-type: none"> <li>• To tell you about the water cycle</li> <li>• I didn't know about a camel, so I tried to read it</li> <li>• I wanted to learn about the first flight</li> </ul>
Learning	The texts was thought to aid learning more generally (generalized to a broader topic or to other learners)	<ul style="list-style-type: none"> <li>• It helps kids learn about cleaning up messes instead of just being all lazy</li> <li>• I like learning about stuff</li> </ul>
Personal connection	Text was related to personal experiences or characteristics	<ul style="list-style-type: none"> <li>• I have a connection because I stay up late</li> <li>• I used to have two [cats] but then my mom gave one to a farm, and then now I have one</li> <li>• Because I have a loose tooth</li> </ul>
Series	The text was part of a preferred series	<ul style="list-style-type: none"> <li>• Because I like reading Frog and toad books</li> <li>• I like Henry and Mudge books</li> </ul>

Social	The text was chosen for social reasons, including recommendations by friends or popularity among peers	<ul style="list-style-type: none"> <li>• P [a classmate] always reads these kind of books, so I just said let me try it</li> <li>• Because I wanted to partner read with D [classmate]</li> </ul>
Story interest	The text had a story that was interesting or enjoyable	<ul style="list-style-type: none"> <li>• It's just so cool that he runs away from some people</li> <li>• Because it was about a girl saving a city and her family</li> </ul>
Style	The text was written in a preferred style (similar to genre)	<ul style="list-style-type: none"> <li>• Because it's violent</li> <li>• It's a scary book</li> <li>• It's a rhyming book</li> <li>• I like spooky books</li> </ul>
Surface	The text had appealing surface features	<ul style="list-style-type: none"> <li>• I really liked the picture</li> <li>• The cover looked cool</li> </ul>
Topic interest	The text was about a favorite topic or an area of interest	<ul style="list-style-type: none"> <li>• Because I like reading about caterpillars and animals</li> <li>• Because I like to learn about dragons</li> <li>• Because I'm really interested in Nickelodeon</li> </ul>
Vocabulary	The text was thought to contain interesting or useful words	<ul style="list-style-type: none"> <li>• It sometimes had funny words</li> <li>• I like the words</li> <li>• It has a lot of good words in it</li> </ul>

## Appendix L: Children's Books Cited

- Aliki. (1988). *How a book is made*. New York: Harper Trophy.
- Ballard, R. D., & Archbold, R. (1998). *Ghost liners: Exploring the world's greatest lost ships*. Illustrated by K. Marschall. Toronto: Madison Press Books.
- Bourgeois, P. (1987). *Franklin in the dark*. Illustrated by B. Clark. New York: Scholastic.
- Brodsky, G. (1990). *The mind of the cat*. New York: Longmeadow Press.
- Brown, R., & Carey, S. (1994). *Hide and seek*. Illustrated by Sal Murdocca. New York: Scholastic.
- Bryant, R. (1995). *Goldilocks and the three bears*. Illustrated by S. Turgeon. Montreal: Tormont Publications.
- Canetti, Y. (1999). *Carlita ropes the twister*. Illustrated by G. B. Karas. Orlando: Steck-Vaughn.
- Canizares, S., & Reed, M. (1998). *Who lives in the rainforest?* New York: Scholastic.
- Cole, J. (1988). *The Magic School Bus at the waterworks*. Illustrated by B. Degen. New York: Scholastic.
- Collicott, S. (1999). *Toestomper and the caterpillars*. New York: Houghton Mifflin.
- Cooper, I. (2000). *Lucy on the loose*. Illustrated by A. Harvey. New York: Random House.
- Curry, J. (2004). *Animal poems*. Illustrated by A. Lewis. London: Scholastic.
- Desimini, L. (2001). *Dot the fire dog*. New York: Scholastic.
- Drake, E., & Steer, D. (2003). *Dragonology: The complete book of dragons*. Cambridge, MA: Candlewick Press.
- Gray, K. (2000). *Eat your peas*. Illustrated by N. Sharratt. New York: Dorling Kindersley Publishing.
- Harrison, J. (1992). *Young people's atlas of the United States*. New York: Kingfisher Books.
- Holub, J. (2003). *Why do horses neigh?* Illustrated by A. DiVito. New York: Puffin.

- James, C. (2000). *My life on an island*. New York: Rosen Publishing Group.
- Jasheway, L. A. (1996). *Bedtime stories for dogs*. Kansas City, MO: Andrews and McMeel.
- McClintock, M. (1958). *A fly went by*. Illustrated by F. Siebel. New York: Random House.
- Nintendo. (2006). *The legend of Zelda: Twilight princess manual*. Redmond, WA: Nintendo of America.
- Osborne, M. P. (2005). *Season of the sandstorms*. Illustrated by S. Murdocca. New York: Random House.
- Page, J. (2000). *Clifford and the big leaf pile*. New York: Scholastic.
- Polacco, P. (1990). *Thunder cake*. New York: Philomel Books.
- Pennypacker, S. (2007). *The talented Clementine*. Illustrated by M. Frazee. New York: Hyperion.
- Rylant, C. (1987). *Henry and Mudge: The first book*. Illustrated by S. Stevenson. New York: Simon and Schuster.
- Scieszka, J. (1996). *The true story of the three little pigs*. Illustrated by L. Smith. New York: Puffin.
- Seuss, Dr. (1968). *The foot book*. New York: Random House.
- Seuss, Dr. (1963). *Hop on pop*. New York: Random House.
- Seuss, Dr. (1960). *Green eggs and ham*. New York: Random House.
- Strasser, T. (1999). *Star Wars Episode I journal: Anakin Skywalker*. New York: Scholastic.

## REFERENCES



## References

- Afflerbach, P. (2007). *Understanding and using reading assessment K-12*. Newark, DE: International Reading Association.
- Allington, R. L. (2006). *What really matters for struggling readers: Designing research based programs*. Boston: Pearson/Allyn & Bacon.
- Anderson, G., Higgins, D., & Wurster, S.R. (1985). Differences in the free-reading books selected by high, average, and low achievers. *The Reading Teacher*, 39(3), 326-330.
- Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly*, 23, 285-303.
- Applegate, M. D., Quinn, K. B., & Applegate, A. J. (2008). *The critical reading inventory: Assessing students' reading and thinking* (2<sup>nd</sup> ed.). Upper Saddle River, NJ: Pearson.
- Armbruster, B. B., Lehr, F., & Osborn, J. (2003). *Put reading first: The research building blocks for teaching children to read*. Jessup, MD: Partnership for Reading, National Institute for Literacy.
- Asher, S.R., Hymel, S., & Wigfield, A. (1978). Influence of topic interest on children's reading comprehension. *Journal of Reading Behavior*, 10(1), 35-47.
- Asher, S.R., & Markell, R.A. (1974). Sex differences in comprehension of high- and low-interest material. *Journal of Educational Psychology*, 66(5), 614-619.
- Atkinson, J. W. (1958). Toward experimental analysis of human motivation in terms of motives, expectancies, and incentives. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society* (pp. 288-305). Princeton, NJ: Van Nostrand.
- Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.
- Atkinson, J. W., & Feather, N. T. (1966). *A theory of achievement motivation*. Huntington, NY: Krieger Publishing Company.
- Atkinson, J. W., & J. O. Raynor (eds.). 1974. *Motivation and achievement*. Washington, DC: Winston.
- Bader, L. A. (1980). *Reading diagnosis and remediation in classroom and clinic*. New York: Macmillan.

- Baker, L., & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly*, 34(4), 452-477.
- Baldwin, R. S., Peleg-Bruckner, Z., & McClintock, A. H. (1985). Effects of topic interest and prior knowledge on reading comprehension. *Reading Research Quarterly*, 20(4), 497-504.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bass, A.S. (2006, December). *A first grade teacher's book selection decisions for emergent readers*. Paper presented at the 56<sup>th</sup> annual meeting of the National Reading Conference, Los Angeles, CA.
- Belloni, L. F., & Jongsma, E. A. (1978). The effects of interest on reading comprehension of low-achieving students. *Journal of reading*, 22(2), 106-109.
- Bernstein, M. R. (1955). Relationship between interest and reading comprehension. *Journal of Educational Research*, 49(4), 283-288.
- Betts, E. A. (1946). *Foundations of reading instruction, with emphasis on differentiated guidance*. New York: American Book Company.
- Brophy, J. (2004). *Motivating students to learn* (2<sup>nd</sup> ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Brown, K. J. (2000). What kind of text – for whom and when? Textual scaffolding for beginning readers. *The Reading Teacher*, 53(4), 292-307.
- Burns, P. C., & Roe, B. D. (1999). *Informal reading inventory: Preprimer to twelfth grade* (5<sup>th</sup> ed.). Boston: Houghton Mifflin Company.
- Caldwell, J. S. (2002). *Reading assessment: A primer for teachers and tutors*. New York: Guilford.
- Carver, R. P. & Leibert, R. E. (1995). The effect of reading library books at different levels of difficulty upon gain in reading ability. *Reading Research Quarterly*, 30(1), 26-48.
- Chall, J. S., & Conard, S. S. (1991). *Should textbooks challenge students? The case for easier or harder books*. New York: Teachers College Press.

- Chapman, J. W., & Tunmer, W. E. (1995). Development of young children's reading self-concept: An examination of emerging subcomponents and their relationship with reading achievement. *Journal of Educational Psychology*, 87(1), 154-167.
- Childress, G.T. (1985). Gender gap in the library: Different choices for girls and boys. *Top of the News*, 42(1), 69-73.
- Clay, M. M. (1985). *The early detection of reading difficulties* (3<sup>rd</sup> ed.). Portsmouth, NH: Heinemann.
- Clay, M. M. (1993a). *An observation survey of early literacy achievement*. Portsmouth, NH: Heinemann.
- Clay, M. M. (1993b). *Reading Recovery: A guidebook for teachers in training*. Portsmouth, NH: Heinemann.
- Cooper, J. L. (1952). The effect of adjustment of basal reading materials on reading achievement. Unpublished doctoral dissertation, Boston University.
- Davis, E. E. (1975). Reading frustration level as indicated by the polygraph. *Journal of Educational Research*, 68(8), 286-288.
- Davis, E. E., & Ekwall, E. E. (1976). Mode of perception and frustration in reading. *Journal of Learning Disabilities*, 9(7), 448-54.
- Davis, Z. T. (1988). A comparison of the effectiveness of sustained silent reading and directed reading activity on students' reading achievement. *The High School Journal*, 72(1), 46-48.
- Deci, E., & Ryan, R. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227-268.
- Dolch, E. (1928). Combined word studies. *Journal of Educational Research*, 17, 11-19.
- Donovan, C.A., Smolkin, L.B., & Lomax, R.G. (2000). Beyond the independent-level text: Considering the reader-text match in first graders' self-selections during recreational reading. *Reading Psychology*, 21(4), 309-333.
- Dzaldov, B. S. & Peterson, S. (2005). Book leveling and readers. *The Reading Teacher*, 59(3), 222-229.
- Eccles, J. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 78-146). San Francisco, CA: W. H. Freeman and Company.

- Ekwall, E. E., Solis, J. K., & Solis, E., Jr. (1973). Investigating informal reading inventory scoring criteria. *Elementary English*, 50(2), 271-274, 323.
- Eldredge, J. L. (1990). Increasing the performance of poor readers in the third grade with a group-assisted strategy. *Journal of Educational Research*, 84(2), 69-77.
- Fielding, L. & Roller, C. (1992). Making difficult books accessible and easy books acceptable. *The Reading Teacher*, 45(9), 678-685.
- Fleming, J. T. (1967). *Children's perception of difficulty in reading materials*. (ERIC Document Reproduction Service No. ED017398)
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.
- Flynt, E. S., & Cooter, R. B., Jr. (2001). *Reading inventory for the classroom* (4<sup>th</sup> ed.). Upper Saddle River, NJ: Merrill Prentice Hall.
- Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading: Good first teaching for all children*. Portsmouth, NH: Heinemann.
- Fountas, I. C., & Pinnell, G. S. (1999). *Matching books to readers: Using leveled books in guided reading, K-3*. Portsmouth, NH: Heinemann.
- Fresch, M. J. (1995). Self-selection of early literacy learners. *The Reading Teacher*, 49(3), 220-227.
- Fry, E. (1977). *Elementary reading instruction*. New York: McGraw Hill.
- Furniss, D. W., & Graves, M. F. (1980). *Effects of stressing oral reading accuracy on comprehension*. Paper presented at the International Reading Association annual convention, St. Louis, MO.
- Gage, N. L., & Berliner, D. C. (1992). *Educational psychology* (5<sup>th</sup> ed.). Boston: Houghton Mifflin.
- Gambrell, L. B., Palmer, B. M., Codling, R. M., & Mazzoni, S. A. (1996). Assessing motivation to read. *The Reading Teacher*, 49(7), 518-533.
- Good, T. L., & Brophy, J. E. (1990). *Educational psychology: A realistic approach* (4<sup>th</sup> ed.). White Plains, NY: Longman.
- Goodman, Y. M., Watson, D. J., & Burke, C. L. (2005). *Reading miscue inventory: From evaluation to instruction* (2<sup>nd</sup> ed.). Katonah, NY: Richard C. Owen Publishers.

- Gough, P. B. (1972). One second of reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press.
- Grote, G. F., & James, L. R. (1991). Testing behavioral consistency and coherence with the Situation-Response Measure of Achievement Motivation. *Multivariate Behavioral Research*, 26, 655-691.
- Guthrie, J. T., Hoa, A. L. W., Wigfield, A., Tonks, S. M., Humenick, N. M., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemporary Educational Psychology*, 32(3), 282-313.
- Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (2004). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading*, 3(3), 231-256.
- Halladay, J. L. (2006, December). *Exploring students' text selection strategies*. Paper presented at the 56<sup>th</sup> annual meeting of the National Reading Conference, Los Angeles, CA.
- Hiebert, E. H. (2005). The effects of text difficulty on second graders' fluency development. *Reading Psychology*, 26(2), 183-209.
- Hiebert, E.H., Mervar, K.B., & Person, D. (1990). Research directions: Children's selection of trade books in libraries and classrooms. *Language Arts*, 67(7), 758-763.
- Hoffman, J. V., Roser, N. L., Salas, R., Patterson, E., & Pennington, J. (2000). *Text leveling and little books in first-grade reading* (Report No. CIERA-R-1-010). Ann Arbor, MI: Center for the Improvement of Early Reading Achievement. (ERIC Document Reproduction Service No. ED439405)
- Hunt, L.C. (1970). The effect of self-selection, interest, and motivation upon independent, instructional, and frustration levels. *Reading Teacher*, 24 (2), 146-151.
- International Reading Association. (1998). *Learning to read and write: Developmentally appropriate practices for young children*. Newark, DE: International Reading Association.
- Johns, J. L. (1997). *Basic reading inventory: Pre-primer through grade twelve and early literacy assessments* (7<sup>th</sup> ed.). Dubuque, IA: Kendall/Hunt Publishing Company.
- Johns, J. L., & Magliari, A. M. (1989). Informal reading inventories: Are the Betts criteria the best criteria?. *Reading Improvement*, 26(2), 124-132.

- Johnson, K. M. (2005). Test review of Gates-MacGinitie Reading Tests(r), Fourth Edition, Forms S and T. From R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook* [Electronic version]. Retrieved August 8, 2008, from the Buros Institute's Test Reviews Online website: <http://www.unl.edu/buros>
- Johnson, M. S., Kress, R. A., & Pikulski, J. J. (1987). *Informal reading inventories* (2<sup>nd</sup> ed.). Newark, DE: International Reading Association.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Jones, M. M., Lignugaris/Kraft, B., & Peterson, S. M. (2007). The relation between task demands and student behavior problems during reading instruction: A case study. *Preventing School Failure*, 51(4), 19-28.
- Jongsma, K. S., & Jongsma, E. A. (1981). Test review: Commercial informal reading inventories. *Reading Teacher*, 34(6), 697-705.
- Juel, C., Griffith, P.L., & Gough, P. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*, 78(4), 243-255.
- Kamil, M. L. (2007, December). *How to get recreational reading to increase reading achievement*. Address presented at the 57th annual meeting of the National Reading Conference, Austin, TX.
- Kasdon, L. M. (1970). Oral versus silent-oral diagnosis. In D. L. DeBoer (Ed.), *Reading Diagnosis and Evaluation: Proceedings of the Thirteenth Annual Convention*. (pp. 86-92). Newark, DE: International Reading Association.
- Kilgallon, P. A. (1942). *A study of relationships among certain pupil adjustments in language situations*. Unpublished doctoral dissertation, Pennsylvania State University.
- Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R.B. Ruddell & N.J. Unrau (Eds.), *Theoretical models and processes of reading* (5<sup>th</sup> ed., pp. 1270-1328). Newark, DE: International Reading Association.
- Kintsch, W., & Franzke, M. (1995). The role of background knowledge in the recall of a news story. In R. F. Lorch, Jr. & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 321-333). Hillsdale, NJ: Erlbaum.
- Kletzien, S. B. (1991). Strategy use by good and poor comprehenders reading expository text of differing levels. *Reading Research Quarterly*, 26(1), 67-86.

- Kragler, S. (2000). Choosing books for reading: An analysis of three types of readers. *Journal of Research in Childhood Education*, 14(2), 113-141.
- Krashen, S. (2005). Is in-school free reading good for children? Why the National Reading Panel is (still) wrong. *Phi Delta Kappan*, 86(6), 444-447.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323.
- Lee, C. D., & Smagorinsky, P. (2000). Introduction: Constructing meaning through collaborative inquiry. In C. D. Lee & P. Smagorinsky (Eds.), *Vygotskian perspectives on literacy research: Constructing meaning through collaborative inquiry*. Cambridge: Cambridge University Press.
- Lennon, C., & Burdick, H. (2004). *The Lexile framework as an approach for reading measurement and success*. Durham, NC: MetaMetrics, Inc.
- Lin, L.-M., Zabucky, K., & Moore, D. (1997). The relations among interest, self-assessed comprehension, and comprehension performance in young adults. *Reading Research and Instruction*, 36(2), 127-139.
- Lipson, M. Y., & Wixson, K. K. (2003). *Assessment and instruction of reading and writing difficulty: An interactive approach* (3<sup>rd</sup> ed.). Boston: Allyn and Bacon.
- Lysaker, J.T. (1997). Learning to read from self-selected texts: The book choices of six first graders. In C.K. Kinzer, K.A. Hinchman, & D.J. Leu (Eds.), *Inquiries in literacy theory and practices: 46<sup>th</sup> yearbook of the National Reading Conference* (pp. 273-282). Chicago: National Reading Conference.
- MacGinitie, W. H., & MacGinitie, R. K. (1988). *Gates-MacGinitie Reading Tests* (3<sup>rd</sup> ed.). Itasca, IL: Riverside Publishing.
- MacGinitie, W. H., & MacGinitie, R. K. (1989). *Technical report: Gates-MacGinitie Reading Tests*. Itasca, IL: Riverside Publishing.
- MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G. & Hughes, (2002). *Gates-MacGinitie Reading Tests* (4<sup>th</sup> ed.). Itasca, IL: Riverside Publishing.
- Mariotti, A. S., & Homan, S. P. (2005). *Linking reading assessment to instruction: An application worktext for elementary classroom teachers* (4<sup>th</sup> ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mathewson, G. C. (1976). The function of attitude in the reading process. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (2nd ed., pp. 655-676). Newark, DE: International Reading Association.

- Mathewson, G. C. (1985). Toward a comprehensive model of affect in the reading process. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (3rd ed., pp. 841-856). Newark, DE: International Reading Association.
- Mathewson, G. C. (1994). Model of attitude influence upon reading and learning to read. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4<sup>th</sup> ed., pp. 1131-1161). Newark, DE: International Reading Association.
- Mesmer, H. A. E. (2006). Beginning reading materials: A national survey of primary teachers' reported uses and beliefs. *Journal of Literacy Research*, 38(4), 389-425.
- Michigan Department of Education, Center for Educational Performance and Information. (2008a). Retrieved June 8, 2008 from [http://www.michigan.gov/documents/cepi/FRLSchFall07\\_232688\\_7.xls](http://www.michigan.gov/documents/cepi/FRLSchFall07_232688_7.xls)
- Michigan Department of Education. (2008b). Retrieved July 3, 2008 from [https://oeaa.state.mi.us/ayp/achievement\\_change\\_2006.asp](https://oeaa.state.mi.us/ayp/achievement_change_2006.asp)
- Mork, T.A. (1973). The ability of children to select reading materials at their own instructional reading level. In W.H. MacGinitie (Ed.), *Assessment Problems in Reading* (pp. 87-95). Newark, DE: International Reading Association.
- Moss, G. & McDonald, J. W. (2004). The borrowers: Library records as unobtrusive measures of children's reading preferences. *Journal of Research in Reading*, 27(4), 401-412.
- Nagy, W.E. & Scott, J.A. (2004). Vocabulary Processes. In R.B. Ruddell & N.J. Unrau (Eds.), *Theoretical models and processes of reading* (5<sup>th</sup> ed., pp. 574-593). Newark, DE: International Reading Association.
- National Center for Education Statistics. (2008). Retrieved July 3, 2008 from [http://nces.ed.gov/globallocator/sch\\_info\\_popup.asp](http://nces.ed.gov/globallocator/sch_info_popup.asp)
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction--Reports of the subgroups*. Washington, DC: National Institute of Child Health and Development.
- O'Connor, R. E., Bell, K. M., Harty, K. R., Larkin, L. K., Sackor, S. M., & Zigmond, N. (2002). Teaching reading to poor readers in the intermediate grades: A comparison of text difficulty. *Journal of Educational Psychology*, 94(3), 474-485.
- O'Hear, M. F., & Ramsey, R. N. (1990). *A comparison of readability scores and student perceptions of reading ease*. (ERIC Document Reproduction Service No. ED321244)



- Olson, A.V. (1984). *Elementary students' self selection of reading material in school libraries*. (ERIC Document Reproduction Service No. ED346082)
- Paris, S. G., & Hoffman, J. V. (2004). Reading assessments in kindergarten through third grade: Findings from the Center for the Improvement of Early Reading Achievement. *The Elementary School Journal*, 105(2), 199-217.
- Pehrsson, R. S. V. (1974). The effects of teacher interference during the process of reading, or how much of a helper is Mr. Gelper? *Journal of Reading*, 17(8), 617-21.
- Pikulski, J. J., & Shanahan, T. (1982). Informal reading inventories: A critical appraisal. In J. J. Pikulski & T. Shanahan (Eds.), *Approaches to the informal evaluation of reading* (pp. 94-116). Newark, DE: International Reading Association.
- Powell, W. R., & Dunkeld, C. G. (1971). Validity of the IRI reading levels. *Elementary English*, 48(6), 637-642.
- Powell, W. R. (1970). Reappraising the criteria for interpreting informal inventories. In D. L. DeBoer (Ed.), *Reading Diagnosis and Evaluation: Proceedings of the Thirteenth Annual Convention*. (pp. 100-109). Newark, DE: International Reading Association.
- Powell, W. R. (1980). Measuring reading performance informally. *Journal of Children and Youth*, 1, 23-31.
- Pressley, M. (2006). *Reading instruction that works: The case for balanced teaching* (3rd ed.). New York: Guilford.
- Pressley, M., Dolezal, S. E., Raphael, L. M., Mohan, L., Roehrig, A. D., & Bogner, K. (2003). *Motivating primary-grade students*. New York: Guilford.
- Pumfrey, P. D. (1985). *Reading: Tests and assessment techniques*, (2<sup>nd</sup> ed.). Kent, UK: Hodder and Stoughton Educational.
- Ramsey, R. N. (1994). Student perception of readability and human interest in upper-level composition textbooks. *Forum for Reading*, 24, 1-10.
- RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D Program in Reading Comprehension*. Santa Monica, CA: RAND.
- Renaissance Learning (2007). *A parent's guide to Accelerated Reader: Questions and answers*. Wisconsin Rapids, WI: Author.

- Rhodes, L. K., & Dudley-Marling, C. (1996). *Readers and writers with a difference: A holistic approach to teaching struggling readers and writers* (2<sup>nd</sup> ed.). Portsmouth, NH: Heinemann.
- Riddle-Buly, M. & Valencia, S.W. (2002). Below the bar: Profiles of students who fail state reading assessments. *Educational Evaluation and Policy Analysis*, 24, 219-239.
- Robb, L. (2000). *Teaching reading in middle school: A strategic approach to teaching reading that improves comprehension and thinking*. New York: Scholastic.
- Rosenblatt, L. M. (1994). The transactional theory of reading and writing. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4<sup>th</sup> ed., pp. 1057-1092). Newark, DE: International Reading Association.
- Rosenblatt, L. M. (1978). *The reader, the text, the poem: The transactional theory of the literary work*. Carbondale, IL: Southern Illinois University Press.
- Ruddell, R. B., & Unrau, N. J. (2004). Reading as a meaning-construction process: The reader, the text, and the teacher. In R.B. Ruddell & N.J. Unrau (Eds.), *Theoretical models and processes of reading* (5<sup>th</sup> ed., pp. 1462-1521). Newark, DE: International Reading Association.
- Rumelhart, D. E. (1994). Toward an interactive model of reading. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4<sup>th</sup> ed., pp. 864-894). Newark, DE: International Reading Association.
- Sagie, A. (1993). Assessing achievement motivation: Construction and application of a new scale using Elizer's multifaceted approach. *The Journal of Psychology*, 128(1), 51-61.
- Schummers, J. L. (1956). Word pronunciation in the oral sight-reading of third grade children. Unpublished doctoral dissertation, University of Minnesota.
- Schunk, D. H., & Zimmerman, B. J. (1997). Developing self-efficacious readers and writers: The role of social and self-regulatory processes. In J. T. Guthrie & A. Wigfield (Eds.), *Reading engagement: Motivating readers through integrated instruction* (pp. 34-50). Newark, DE: International Reading Association.
- Smith, L. L., & Joyner, C. R. (1990). Comparing recreational reading levels with reading levels from an informal reading inventory. *Reading Horizons*, 30(4), 293-299.
- Spache, G. (1963). *Diagnostic reading scales*. Monterey: California Test Bureau.
- Spence, J. T., & Helmreich, R. L. (1983). Achievement-related motives and behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and*

- sociological approaches* (pp. 10-74). San Francisco, CA: W. H. Freeman and Company.
- Stahl, S. A. (2004). What do we know about fluency?: Findings of the National Reading Panel. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 187-211). Baltimore: Paul H. Brookes Publishing.
- Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360-407.
- Stieglitz, E. L. (2002). *The Stieglitz informal reading inventory: Assessing reading behaviors from emergent to advanced levels* (3<sup>rd</sup> ed.). Boston: Allyn and Bacon.
- Szymusiak, K., & Sibberson, F. (2001). *Beyond leveled books: Supporting transitional readers in grades 2-5*. Portland, ME: Stenhouse.
- Taylor, B.M., Frye, B.J., & Maruyama, G.M. (1990). Time spent reading and reading growth. *American Educational Research Journal*, 27(2), 351-362.
- Thorndike, E. (1921). Word knowledge in the elementary school. *Teachers College Record*, 22, 334-370.
- Timion, C. S. (1992). Children's book selection strategies. In J. W. Irwin & M. A. Doyle (Eds.), *Reading/writing connections: Learning from research* (pp. 204-222). Newark, DE: International Reading Association.
- Tompkins, G. E. (2006). *Literacy for the 21<sup>st</sup> century: A balanced approach* (4<sup>th</sup> ed.). Upper Saddle River, NJ: Pearson.
- Valmont, W. J. (1972). Creating questions for informal reading inventories. *The Reading Teacher*, 25(6), 509-512.
- Vaughn, B. J., & Horner, R. H. (1997). Identifying instructional tasks that occasion problem behaviors and assessing the effects of student versus teacher choice among these tasks. *Journal of Applied Behavior Analysis*, 30, 299-312.
- Voss, J. F., & Silfies, L. N. (1996). Learning from history text: The interaction of knowledge and comprehension skill with text structure. *Cognition and Instruction*, 14, 45-68.
- Vygotsky, L.S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Weiner, B. (1972). *Theories of motivation: From mechanism to cognition*. Chicago: Markham.

- White, R., & Jordan, W. (1987). *Readability study*. Albuquerque, NM: Albuquerque Technical Vocational Institute. (ERIC Document Reproduction Service No. ED284043)
- Wigfield, A., & Eccles, J. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68-81.
- Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., et al. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools*, 45(5), 432-445.
- Wolfe, M. B. W., & Mienko, J. A. (2007). Learning and memory of factual content from narrative and expository text. *British Journal of Educational Psychology*, 77(3), 541-564.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Child Psychiatry*, 17, 89-100.
- Woods, M. L., & Moe, A. J. (2007). *Analytical reading inventory: Comprehensive standards-based assessment for all students, including gifted and remedial* (8<sup>th</sup> ed.). Upper Saddle River, NJ: Pearson.
- Worthy, J., Sailors, M. (2001). "That book isn't on my level": Moving beyond text difficulty in personalizing reading choices. *The New Advocate*, 14(3), 229-239.
- Zahorik, J. A. (1987). Reacting. In M. J. Dunkin (Ed.), *International encyclopedia of teaching and teacher education* (pp. 416-423). Oxford: Pergamon Press.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02956 9377