

EVALUATING PASSAGE-LEVEL CONTRIBUTORS
TO TEXT COMPLEXITY

by

SHAHEEN MUNIR-MCHILL

A DISSERTATION

Presented to the Department of Special Education
and Clinical Sciences
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2013

DISSERTATION APPROVAL PAGE

Student: Shaheen Munir-McHill

Title: Evaluating Passage-Level Contributors to Text Complexity

This dissertation has been accepted and approved in partial fulfillments of the requirements for the Doctor of Philosophy degree in the Department of Special Education and Clinical Sciences by:

Roland H. Good, III	Chair
Kelli C. Cummings	Core Member
Elizabeth Harn	Core Member
Laura Lee McIntyre	Core Member
Gina Biancarosa	Institutional Representative

and

Kimberly Andrews Espy	Vice President for Research & Innovation/Dean of the Graduate School
-----------------------	--

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded September 2013

© 2013 Shaheen Munir-McHill

DISSERTATION ABSTRACT

Shaheen Munir-McHill

Doctor of Philosophy

Department of Special Education and Clinical Sciences

September 2013

Title: Evaluating Passage-Level Contributors to Text Complexity

The complexity of text has a number of implications for educators in the areas of instruction and assessment. Text complexity is particularly important in formative assessments, which utilize repeated, alternate, equivalent forms to capture student growth towards a general outcome. A key assumption of such tools is that alternate forms of the assessment are of equal complexity. Consequently, there is a need to better understand what variables contribute to text complexity and how they impact student performance. This study was designed to evaluate features of text that are not typically included in readability estimates but may contribute to the text complexity: text cohesion and genre.

Currently, text complexity of oral reading fluency measures is often quantified using readability estimates. It is hypothesized that a factor generally excluded from readability estimates, text cohesion – the extent to which the text functions as a cohesive, meaningful whole – contributes to text variability and variability in student performance. This research evaluated the role of a type of text cohesion (referential cohesion) in text complexity by manipulating the cohesion of passages otherwise assumed to be of equal difficulty. Genre was also considered, as research suggests that genre may impact complexity ratings of texts. Passages were strategically selecting to capture four conditions – 1) informational text/low cohesion, 2) informational text/high cohesion, 3)

narrative text/low cohesion, and 4) narrative text/high cohesion. Data were collected on reading rate, accuracy, and passage-specific reading comprehension

Results were analyzed using two-way, univariate ANOVA with dependent observations. Results indicate effects for each of the dependent variables included in the design. For rate and accuracy, results indicate significant interactions between genre and referential cohesion; scores were significantly higher for high cohesion narrative text than low cohesion narrative text and high cohesion informational text. There was a significant main effect of genre on comprehension, with students performing significantly better on the comprehension measure for narrative texts than informational texts. Altogether, these results indicate direct effects of genre and referential cohesion on student reading performance and provide evidence that text cohesion may be a meaningful component of text complexity.

CURRICULUM VITAE

NAME OF AUTHOR: Shaheen Munir-McHill

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of Southern California, Los Angeles

DEGREES AWARDED:

Doctor of Philosophy, School Psychology, 2013, University of Oregon
Master of Science, Special Education, 2011, University of Oregon
Bachelor of Arts, Psychology, 2006, University of Southern California

AREAS OF SPECIAL INTEREST:

Data-Based Decision Making
Formative Assessment in Reading

PROFESSIONAL EXPERIENCE:

School psychology intern, Springfield Public Schools, Springfield, OR, 2012-2013

Guest lecturer, Department of Education Studies, University of Oregon, 2011

Guest lecturer, Department of Special Education and Clinical Services, University of Oregon, 2011

Supervised college teaching assistant, Department of Special Education and Clinical Services, University of Oregon, 2010-2011

GRANTS, AWARDS, AND HONORS:

College of Education Doctoral Research Grant, University of Oregon, 2013

DIBELS Student Support Award, University of Oregon, 2008, 2010

Graduate Teaching Fellowship, Department of Special Education and Clinical

Services, 2008-2012

Segal Americorps Education Award, 2007

Summa cum Laude, University of Southern California, 2006

ACKNOWLEDGEMENTS

I would like to extend my deepest gratitude to the following people:

My committee chair and advisor, Dr. Roland Good, III, for his infectious enthusiasm for the work and unwavering belief in my skills as a researcher, which kept me going long after the last piece of chocolate was gone;

The members of the dissertation committee, Dr. Gina Biancarosa, Dr. Kelli Cummings, Dr. Elizabeth Harn, and Dr. Laura Lee McIntyre, for challenging me to never settle for mediocre when I am capable of so much more;

The University of Oregon and the Dynamic Measurement Group for financial and emotional support of this project;

The graduate student data collectors who volunteered their time and skills to ensure that this research could happen: Emily Barrett, Vincent Campbell, Kelly Collins, Ronda Fritz, and Caitlin Rasplica;

And the cooperating schools and teachers, for welcoming me into their classrooms and reminding me why our work is so important.

TABLE OF CONTENTS

Chapter	Page
I. STATEMENT OF THE PROBLEM	1
Text Complexity	2
The Importance of Text Complexity in Instruction and Assessment	3
College and Career Readiness	4
School Accountability.....	5
Alternate Forms in Formative Assessment.....	6
Alternate Forms for Evaluating Metrics of Text Complexity.....	8
Text Complexity Metrics	9
Reader and Task Considerations.....	9
Quantitative Dimensions.....	10
Qualitative Dimensions.....	12
Evaluating Qualitative Dimensions of Text.....	12
Text Cohesion	13
Integrated Model of Cohesion	15
Grammatical Cohesion.....	15
Syntactic Structure	15
Narrative Structure.....	18
Lexical Cohesion	21
Lexical Accessibility and Diversity	21
Referential Cohesion.....	22
Selection of Referential Cohesion Index for Study	24
Passage Genre	25
Purpose of This Research and Hypotheses	26
Direct Effects on Oral Reading Fluency Rate.....	27
Direct Effects on Oral Reading Fluency Accuracy.....	27
Direct Effects on Passage-Specific Comprehension.....	28
Research Questions	28
II. LITERATURE REVIEW	30
What Makes Text Difficult?	30
Text-Based Features of Text Complexity.....	30
Decoding Difficulty	31
Semantic Difficulty.....	31

Chapter	Page
Syntactic Difficulty	32
Coherence and Cohesion.....	32
Reader-Based Features of Text Complexity	33
Approaches to Evaluating Text Complexity.....	33
Readability Formulas	34
2009 NAEP Reading Framework	38
Common Core Standards Framework.....	39
Text Cohesion: A Potential Contribution to the Evaluation of Text Complexity.....	43
Effects of Cohesion as a Whole	43
Effects of Referential Cohesion	45
Cohesion and Readability: Related but Distinct Constructs	48
Interactions Between Cohesion and Genre	48
Quantifying Text Cohesion Using Coh-Metrix	50
Summary.....	53
III. METHODOLOGY	55
Independent Variables	55
Genre.....	55
Referential Cohesion.....	56
Adjacent Anaphor Overlap	58
Adjacent Argument Overlap.....	58
Content Word Overlap.....	59
Stem Overlap	59
Latent Semantic Analysis (Sentence All)	61
Constructing the RCCS.....	61
Readability	63
Manipulating Independent Variables: Passage Selection	64
Measure Referential Cohesion and Identify “High” and “Low” Cohesion Passages	65
Identify Passages with Similar Readability Scores.....	65
Identify Two Passages Within Each Genre	66
Dependent Variables	66
Dependent Variable #1: Rate	67
Dependent Variable #2: Accuracy	67
Dependent Variable #3: Comprehension	68

Chapter	Page
Measures	69
Oral Reading Fluency	69
Passage Recall.....	72
Conservative	74
Liberal	74
No Match-Consistent	75
No Match-Inconsistent.....	75
Participants.....	78
Procedure	79
Data Collector Training	79
Data Collection	79
Coding of Passage Recalls	80
Participant Incentives.....	81
Summary.....	81
IV. RESULTS.....	83
Characteristics of the Invited Sample	84
Characteristics of the Actual Sample	85
Data Transformations.....	87
Descriptive Statistics.....	87
Intercorrelations	88
Oral Reading Fluency Rate.....	91
Oral Reading Fluency Accuracy.....	93
Passage-Specific Reading Comprehension.....	96
Summary.....	97
V. CONCLUSION	99
Discussion.....	99
Interpretations of Non-Significant Relation Between Referential Cohesion and Comprehension	99
Potential Effects of Background Knowledge.....	102
Cohesion and Grade Level.....	103
Implications.....	103
Implications for Instruction.....	103
Implications for Curriculum-Based Measurement.....	104

Chapter	Page
Implications for Measurement of Text Complexity.....	107
Study Limitations.....	109
Next Steps.....	112
Replication.....	112
Future Directions for Measurement of Referential Cohesion.....	112
Future Directions in Measurement of Comprehension.....	114
Summary.....	116
REFERENCES CITED.....	119

LIST OF FIGURES

Figure	Page
1. Graph of student progress monitoring data illustrating “bounce” in student performance	9
2. Integrated model of text cohesion.....	16
3. A model of relations between independent variables and dependent variables.....	27
4. Measurement model of referential cohesion.....	58
5. Pairwise comparisons of interaction effects between referential cohesion and genre on oral reading fluency rate	93
6. Pairwise comparisons of interaction effects between referential cohesion and genre on oral reading fluency accuracy	95
7. Main effect of genre on passage-specific reading comprehension	97
8. Revisited model of relations between independent and dependent variables.....	100

LIST OF TABLES

Table	Page
1. Qualitative Dimensions of Text Complexity Included in the Common Core Standards Framework.	41
2. Results of Selected Studies Evaluating the Effects of Revisions to Improve Referential Cohesion on Reading Comprehension Performance.....	49
3. Expert Reviewer and Passage Author Judgments of Passage Genre.....	57
4. Coh-Metrix Variables Included in the Researcher-Developed Referential Cohesion Composite Score (RCCS).....	60
5. Inter-Correlations Between Variables Included in the Referential Cohesion Composite Score (Z-Scores).....	62
6. DIBELS Next Oral Reading Fluency Passages Selected for Study Inclusion.....	66
7. Descriptive Statistics for easyCBM Benchmark and Study Passage Rate Scores (First Minute and Pro-Rated Whole Passage).....	72
8. Correlations Between easyCBM Benchmark and Study Passage Rate Scores (First Minute and Pro-Rated Whole Passage).....	73
9. Sample Coding of Student Responses to the Passage Retell Task	76
10. Descriptive Statistics for Rate (Pro-Rated Whole Passage), Accuracy, and Comprehension for all Passages Included in Study	88
11. Descriptive Statistics by Risk Level for Rate (Pro-Rated Whole Passage), Accuracy, and Comprehension for all Passages Included in Study.....	89
12. Intercorrelations Between Oral Reading Fluency Rate, Oral Reading Fluency Accuracy, and Passage-Specific Comprehension Scores for All Measures	90
13. Two-Way, Within-Subjects Analysis of Variance Summary Table for the Effect of Genre and Cohesion on Oral Reading Fluency Rate.....	91
14. Two-Way, Within-Subjects Analysis of Variance Summary Table for the Effect of Genre and Cohesion on Oral Reading Fluency Accuracy	94
15. Two-Way, Within-Subjects Analysis of Variance Summary Table for the Effect of Genre and Cohesion on Passage-Specific Reading Comprehension.....	96

CHAPTER I

STATEMENT OF THE PROBLEM

Every day, educators face a multitude of questions about the complexity of text. Teachers aiming to match students to text may wonder: how can I assess the difficulty of this book? How do I determine if it is aligned with my student's skills and needs? How can I be systematic in the assignment of reading materials, so that the demands placed on the student grow commensurate with the student's skills? Interpreters of test results may ask themselves: how challenging is the text in this assessment? How comparable are alternate forms of this assessment in terms of difficulty? How does the complexity of this assessment align with course content? These, among other questions, highlight the critical role of text complexity in teaching and assessment.

Text complexity is defined by the Common Core Standards in English and Language Arts (2010) as the “inherent difficulty of reading and comprehending a text combined with consideration of reader variables” (Glossary, p. 43). This definition suggests that there are characteristics about the text itself that interacts with reader features to determine the complexity of a given text. While reader variables are a critical component of this definition, it highlights that texts contain “inherent difficulty” that is independent of reader variables. The purpose of this research is to evaluate those “inherent” features of text that contribute to complexity.

In this chapter, the background and importance of the study will be outlined. First, the components of text complexity will be defined and described, including decoding difficulty, semantic difficulty, syntactic complexity, genre, and especially text cohesion, the extent to which a text hangs together to form a coherent whole. Second, the

importance of text complexity in instruction as well as in assessment will be described, with particular focus on the role of text complexity in formative assessment. Third, considerations in evaluating text complexity are described, including: 1) consideration of the reader and task, 2) quantitative measures, and 3) qualitative dimensions. In the fourth section, a next step for improving estimates of text complexity by quantifying traditionally qualitative features of text will be proposed. Specifically, it is proposed that text cohesion can be quantified, and should be included along with passage genre in measures of text complexity. Because cohesion can be described and disaggregated in many different ways, the model underlying the use of the term cohesion in this context will be described. Finally, this chapter will describe how a measure of text cohesion along with estimates of genre may be used to evaluate text complexity of formative assessment tools.

Text Complexity

Text complexity refers to the text-derived difficulty of a given passage. While reader-based factors such as background knowledge also contribute to the difficulty of the passage, this discussion of text complexity focuses on those features central to the text itself that make the passage more or less challenging to decode and understand. Many components are involved in text complexity, including decoding difficulty, semantic difficulty, syntactic difficulty, genre, and text cohesion, among others. Decoding difficulty describes the decoding demands placed on the student, sometimes measured by the average length of words in the passage or the average number of syllables per word. Semantic difficulty captures the semantic requirements of the text, especially the familiarity and difficulty of the vocabulary in the text. Syntactic difficulty

refers to the role of the syntactic structure of the text in supporting reader decoding and understanding, both at the sentence level (e.g., variability of sentence structure) and at the global level (e.g., passage flow and organization). According to the Florida Center for Reading Research (2006), genre refers to “different styles of text that reflect a variety of purposes which children encounter when reading” (e.g., narrative, informational). While many classifications exist, research on text complexity generally focuses on the differential demands of narrative or prose versus informational or expository text. Finally, text cohesion describes the extent to which the text hangs together to form a coherent whole. Cohesive texts provide appropriate linkages between ideas, concepts, narrative elements (e.g., time, setting, characters), and themes to support reader comprehension of the text. This list of text complexity elements is not comprehensive; other text-based features may contribute to the difficulty of the passage, such as the complexity of the ideas and concepts expressed; however, the described elements can be more readily operationalized and measured and, as a result, are the major areas explored in the literature.

The Importance of Text Complexity in Instruction and Assessment

Understanding and capturing the components that contribute to text complexity has implications for both instruction and assessment. Instructionally, the Common Core Standards Initiative (2010) stress that students develop skills to be able to read and comprehend texts of increasing complexity as they progress through school. This expectation is based on data documenting the importance of comprehending complex texts in college and the workplace. In assessment, knowledge and understanding of text complexity has implications for both summative and formative assessment. For

summative assessments such as state accountability tests, understanding of text complexity may help to improve test construction and interpretation. For formative assessment, controlling text complexity is critical in facilitating accurate individual decisions. Additionally, improved measures of text complexity will facilitate the development of better progress monitoring materials. This section describes the role of text complexity in the instruction and assessment domains, and builds a case for better understanding features that contribute to text complexity.

College and career readiness. According to a 2006 report by college readiness test developer ACT, Inc., the ability to answer questions about complex texts appears to differentiate between students who achieve the benchmark on the ACT reading test and those that do not (ACT, 2006). As described in the report, the complexity of all reading passages was ranked on a three-point qualitative ranking scale, and performance on those passages ranked as “complex” was the clearest differentiator over inference making and cognitive skills such as identifying the main idea or the meanings of words in context. In fact, students performing below the benchmark performed no better than chance on these test items, highlighting this skill’s impact on overall reading proficiency. These findings were consistent across gender, race/ethnicity, and socio-economic status.

Additionally, there is evidence that college and workplace texts are significantly more complex than high school texts, and that this discrepancy in text complexity has increased over time (Common Core Standards Initiative Appendix A, 2010). Evaluations of the complexity of K-12 school reading materials indicate that complexity demands have steadily decreased on measures of readability and vocabulary since the middle of the 19th century. Students are also provided with more scaffolding and support in reading

school texts, decreasing independent reading demands. At the same time, the complexity of college and career reading materials has increased, with increasing emphasis on informational texts like periodicals and independent reading. This discrepancy in K-12 and college/career reading demands indicates that students graduating from high school may be unprepared for the reading demands of college and the work force. Accordingly, educators need accurate measures of text complexity to 1) identify target levels of complexity students should attain, and 2) provide systematic increases in complexity by grade to attain those standards.

In order to achieve these goals, researchers and educators must better understand what contributes to text complexity and how to teach students strategies to understand complex text. First, educators must understand the features of texts that impact student comprehension. Only then can educators prepare students for college and career reading demands by sequencing and ordering text in systematic steps of increasing complexity so that students develop skills on less complex text early, develop skills with text of increasing complexity in elementary and middle school, and are able to engage with text of high complexity linked to college and career readiness in high school. Consequently, an evaluation of text complexity factors is a critical prerequisite to building skill in text with increasing complexity and improving understand of texts that may lack inherently supportive text structures.

School accountability. As a result of legislation such as No Child Left Behind, summative assessment data are being used to make decisions about school effectiveness. These decisions have potentially serious implications for school funding and resource allocation. While each state has its own accountability assessment, these results are

being used nationally to interpret state performance in reading and content areas. However, state assessments are not designed to be comparable in difficulty. Consequently, student performance may differ as a function of the assessment, rather than state instructional practices or student performance. Capturing text complexity information about the assessments could alleviate some of these concerns in two ways. First, standards of text complexity could be used during test development, so that assessments are written within a given band of text complexity for each grade level assessed. This would allow for better understanding of the demands of each assessment and comparability across states. Second, tests may be evaluated on the basis of text complexity post hoc to better interpret student performance. For example, clear operationalizations of text complexity designations would allow evaluators to better understand what skills the assessments are measuring, and how students are performing compared to those skills.

Alternate forms in formative assessment. Accurate measurement of text complexity plays a particularly important role in general outcome measurement, where student performance is monitored over time using a common metric and criterion. This type of assessment, called formative assessment, is used to inform instructional practices and facilitate decision-making. Unlike summative assessments, which capture student achievement at the conclusion of an instructional unit, formative assessments allow educators to evaluate student learning as it is occurring and adjust instruction in response to student needs. This is accomplished through the administration of repeated alternate equivalent forms, which capture student growth towards a general outcome or goal.

Formative assessment is made possible due the development of assessment tools

like curriculum-based measurement (CBM). CBM is an approach to assessing student progress towards critical skills. Unlike mastery measurements, which capture discreet skill mastery, goal oriented monitoring of basic early literacy skills allows CBM to monitor the development of skills toward a meaningful outcome (Deno, 2003). An essential component in CBM is its repeatability; CBM is designed to capture growth over time. Performance can be plotted on an individual student graph, to allow for evaluation of past, present, and projected rates of growth.

Within this CBM framework, educators can make decisions about student performance by comparing expected and actual rates of progress. By plotting baseline performance and a goal, educators can create an aimline illustrating the rate of progress necessary to reach the goal. Student performance data are then plotted and compared to the aimline, in order to make decisions about student progress (Deno & Marston, 2006). Consistent performance below the aimline indicates a need to adjust support, while performance at or above the aimline indicates a strong likelihood that the student will achieve the desired level of performance.

A key assumption of formative assessment measures like CBM is that all alternate forms of the assessment are of equal complexity. Equivalent forms of CBM probes allow changes in student performance to be attributed to student growth rather than probe effects. Consequently, form equivalency of CBM probes is critical in making valid and reliable decisions about student performance. Passage equivalency in text complexity is thus a major component of CBM probe development and selection. Test developers may attempt to control passages to be of uniform complexity though a variety

of means, such as expert review, targeted readability criteria, and pilot testing (Albano & Rodriguez, 2012).

Passages of relatively uniform text complexity are also used to evaluate student change in response to instruction. With effective reading instruction, student performance should increase as a result of growth in student skill. In practice, however, student performance is not always so consistent. Students may display “bounce,” or inconsistent performance, across multiple progress monitoring points (see Figure 1). Such variability presents challenges for decision-making, as estimates of the student’s true level of skill and progress are clouded by inconsistent performance (Parker, Vannest, Davis, & Clemens, 2010). There are at least three types of factors that may contribute to such variability in student performance: passage-level factors (e.g., readability, genre, and cohesion), student-level factors (e.g., background knowledge and interest), and environmental factors (e.g., testing conditions and familiarity of tester). While student-level and environmental factors may be challenging to control, it is important to evaluate means of reducing variability due to passage-level factors in order to reduce bounce as much as possible.

Alternate forms for evaluating metrics of text complexity. Because uniformity of text complexity is a key assumption of CBM, alternate forms provide an opportunity to evaluate formulas for capturing text complexity. Specifically, CBM offers a technology to evaluate relations between measures of text complexity and student performance. For example, CBM passages written to adhere to strict standards of readability allow researchers to control readability to evaluate the effects of other measures that may contribute to text complexity. The unique nature of CBM alternate forms – far more

alternate forms than a published norm referenced test – allows this technology to be used in ways that other assessments cannot in order to evaluate those indices of text complexity.

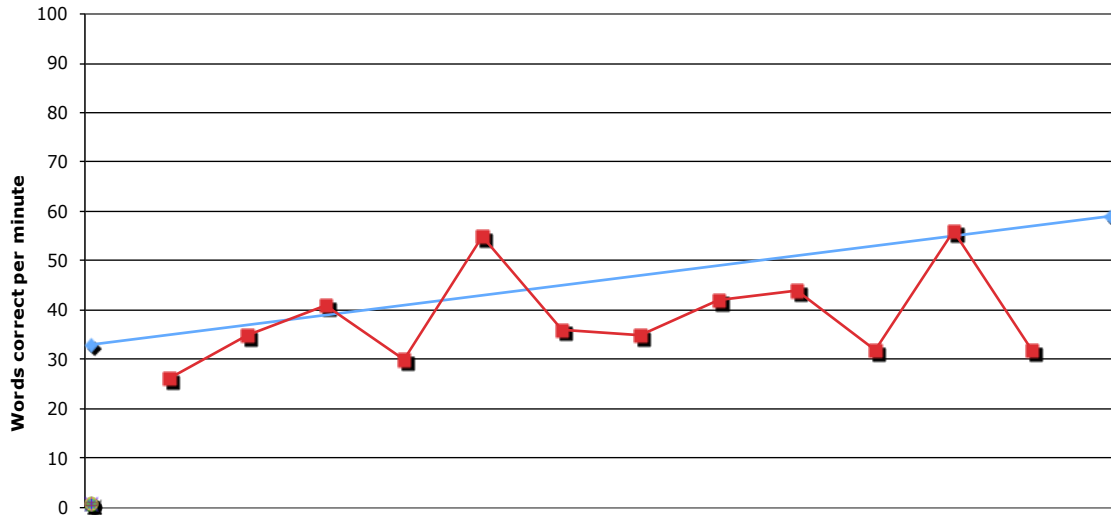


Figure 1. Graph of student progress monitoring data illustrating “bounce” in student performance.

Text Complexity Metrics

The Common Core Standards (2010) propose a three-part model for measuring text complexity – reader and task considerations, quantitative dimensions, and qualitative dimensions. Authors of the Common Core Standards suggest that estimates of text complexity include evaluation of all three of these domains. A description of each of these domains as well as considerations for evaluation are described below.

Reader and task considerations. Reader and task considerations capture the features of text complexity that are individual to the student and the environment of the reading task. These considerations include individual reader characteristics, such as background knowledge, decoding skills, and comprehension strategies. They also

capture features of the environment that may impact performance, such as day of the week, time of day, and environmental stimuli like the presence or absence of other students. Finally, these considerations capture the interaction between the reader and the environment. For example, a student's individual reading skills may impact performance before lunch or when there are multiple activities occurring in the classroom, but not in other environments or contexts.

While these considerations likely contribute to the complexity of a reading task for a given child in a given environment, these features are challenging to capture. Because student and task considerations vary by individual student and context by definition, it is difficult to evaluate or control differences in text complexity due to student and task factors.

Quantitative dimensions. Quantitative measures capture the features of text complexity that can be quantified and counted. Quantitative features are generally included in readability formulas. Readability formulas are based on readily observable, countable features of the text that are generally organized into three factors – decoding difficulty, semantic difficulty, and syntactic difficulty. Decoding difficulty describes the demands of student decoding skill. Decoding difficulty is not included in all readability formulas (such as the Lexile Framework for Reading), but can be quantified by counting the number of characters per word or the number of syllables per word (e.g., Powell-Smith, Good, & Atkins, 2010). Semantic difficulty addresses the semantic requirements of the words in the text, such as the familiarity or uniqueness of the vocabulary used. For example, in the Lexile Framework, semantic difficulty is quantified by counting the mean log of word frequency based on a corpus of approximately 600 million words (Lennon &

Burdick, 2004). Syntactic difficulty captures features of grammar and sentence structure. For example, in the Lexile Framework, syntactic difficulty is quantified as the mean length of sentences in the text.

While these counts capture some of the passage-level contributors to text complexity, readability formulas only consider surface-level features of the text and may fail to capture other features that affect the comprehensibility of the text (e.g., Foorman, 2009; Hiebert, 2011). In some studies, readability formulas have been found to account for some of the variability in text (e.g., Briggs, 2011), but research has consistently found substantial variability in student performance that is unexplained by readability formulas (e.g., Francis, Santi, Barr, Fletcher, Varisco, & Foorman, 2008; Ardoin, Williams, Christ, Klubnik, & Wellborn, 2010). While some of the unexplained variability may be attributed to the student and task considerations, it is possible that there may be other text-level features of text complexity to consider that may be controllable.

One approach to improving readability formulas as measures of text-level text complexity is to attempt to quantify some of the qualitative features of text complexity. For example, the overlap of content and structure may be captured by counting the proportion of sentences that contain overlapping content words, nouns, arguments, or sentence stems. Similarly, the conceptual overlap of words in the text can be evaluated by using latent semantic analysis to generate a quantitative measure of semantic relations. The Coh-Metrix program (McNamara, Louwerse, Cai, & Graesser, 2005) is designed to provide a quantifiable metric that may correspond with some of these qualitative features of text complexity. The Coh-Metrix program provides quantitative counts of a number of

components of text cohesion, and may potentially be one tool to aid researchers in evaluating the qualitative dimensions of text.

Qualitative dimensions. Qualitative features of text complexity are features of the text that may affect a reader understanding of text, but that may be challenging to quantify. The Common Core Standards (2010) describe four qualitative factors that may impact text complexity: 1) the levels of meaning or author's purpose for writing, 2) overall structure and format of the text, 3) use of language in conventional vs. unconventional ways and clarity of the language used, and 4) the background knowledge demands of the text. Additionally, other features such as the complexity of ideas or author's message may impact the demands placed on the reader. Unlike quantitative features of text, the Common Core Standards suggest that qualitative features of text are best evaluated through expert judgment and discussion; however, it may be possible to capture some traditionally qualitative features of text complexity through quantitative analysis.

Evaluating Qualitative Dimensions of Text

As noted above, quantitative estimates of text complexity may benefit from the inclusion of some features of text typically reserved for qualitative analysis. While the quantitative measurement of traditionally qualitative features of text will not replace qualitative analysis, it may help improve estimates of quantitative complexity by including a broader range of the variables that impact text complexity. In particular, it may be possible to quantify some aspects of the overall structure and format of the text, one of the four qualitative factors identified by the Common Core Standards. Two

components of the overall text structure and format –text cohesion and passage genre – may be suitable for such an analysis.

Text Cohesion

One feature of text structure and format that may be amenable to being quantified and incorporated into estimates of text complexity is text cohesion. Text cohesion describes the extent to which text hangs together to form a coherent whole (Morris & Hirst, 1991). This definition suggests a qualitative assessment of the overall structure and clarity of the text in delivering the intended message. However, text cohesion can be disaggregated into component parts (cohesive devices), which may support understanding of the qualitative components of text complexity. Evaluation of these devices may be one means of understanding and evaluating the qualitative features that contribute to text complexity and the linkages between the qualitative and quantitative dimensions of text.

Text cohesion can be described as the extent to which a passage constitutes a unified whole. Specifically, cohesion captures the ties between idea units within a text, and is what differentiates a passage from a series of sentences. Take, for example, the following sentences:

The nation of Fiji is made up of more than 300 individual islands.

The almost century-long occupation by the British ended in 1970.

The sugar in sugarcane is extracted with water or by diffusion.

As they stand, there is little to nothing connecting these sentences together to form a meaningful whole. While an experienced reader may use background knowledge or inferencing skills to attempt to create meaningful connections between the sentences (e.g., by using known information about Fiji’s colonization by the British to infer that the

“century-long occupation” refers to Britain’s occupation of Fiji), such connections are not supported by the text. Cohesion may be imposed upon these sentences by creating connections between ideas, such as:

The nation of Fiji is made up of more than 300 individual islands. Fiji was occupied by the British for almost a century, but this occupation came to an end in 1970. One of the primary industries of Fiji is sugar processing, which requires sugar from sugarcane to be extracted with water or by diffusion.

In this example, the sentences are connected by a number of devices, such as the repetition of the key word “Fiji” in all three sentences and the use of conjunctions like “but” and “which.” Unlike the first example, these sentences now constitute a cohesive text.

While the latter example demonstrates how cohesion connects sentences into meaningful texts, it could certainly be re-written to make such connections even more explicit. In doing so, the text would represent a greater degree of cohesion than either of the provided examples. Texts may vary in the degree of cohesion because cohesion exists upon a continuum; the presence or absence of ties between items in the text and across the entire text affect the cohesiveness of the selected passage.

In order to better understand what makes a text cohesive, researchers have examined the construct of cohesion in a variety of different ways (e.g., Halliday & Hasan, 1976; Kintsch & van Dijk, 1978). The Halliday and Hasan model of cohesion defines cohesion as the semantic relations within the text, which can be coded at three levels: 1) the semantic system, 2) the lexicogrammatical system, and 3) the phonological and orthographic systems. This model focuses heavily on devices within a text that promote

cohesion – co-reference, substitutions, ellipses, conjunctions, and reiteration – and does not emphasize how these devices interact with the reader to impact interpretation or understanding of the text. In contrast, the Kintsch and van Dijk model focuses primarily on the interactions between the surface-level textbase (e.g., many of the devices described by Halliday and Hasan), the meaning of the passage, and the reader. This model emphasizes cohesion as a means of supporting the reader’s construction of text meaning. Each of these models provides important contributions to the field’s understanding of cohesion; rather than selecting one existing model of cohesion, this paper presents a synthesis of different approaches to evaluating cohesion, which will be called the integrated model of cohesion.

Integrated model of cohesion. In the integrated model of cohesion, cohesion is conceptualized in two distinct ways – grammatical cohesion and lexical cohesion. Each type of cohesion as well as the elements and devices within grammatical and lexical cohesion are described below. The entire model is summarized in Figure 2.

Grammatical cohesion. Grammatical cohesion refers to ties and connections between elements of the text that are grammatical in nature (Halliday & Hasan, 1976). One component of grammatical cohesion is the redundancy and complexity of sentence structure, called the syntactic structure. Additionally, grammatical cohesion includes the use of a predictable structure across the text – such as the logical inclusion of predictable elements of story grammar – and the maintenance of consistency of space and time, which can be described as the narrative structure of the text.

Syntactic structure. The syntactic structure of a text captures the variability of sentence structures across the text as well as the complexity of such sentence structures

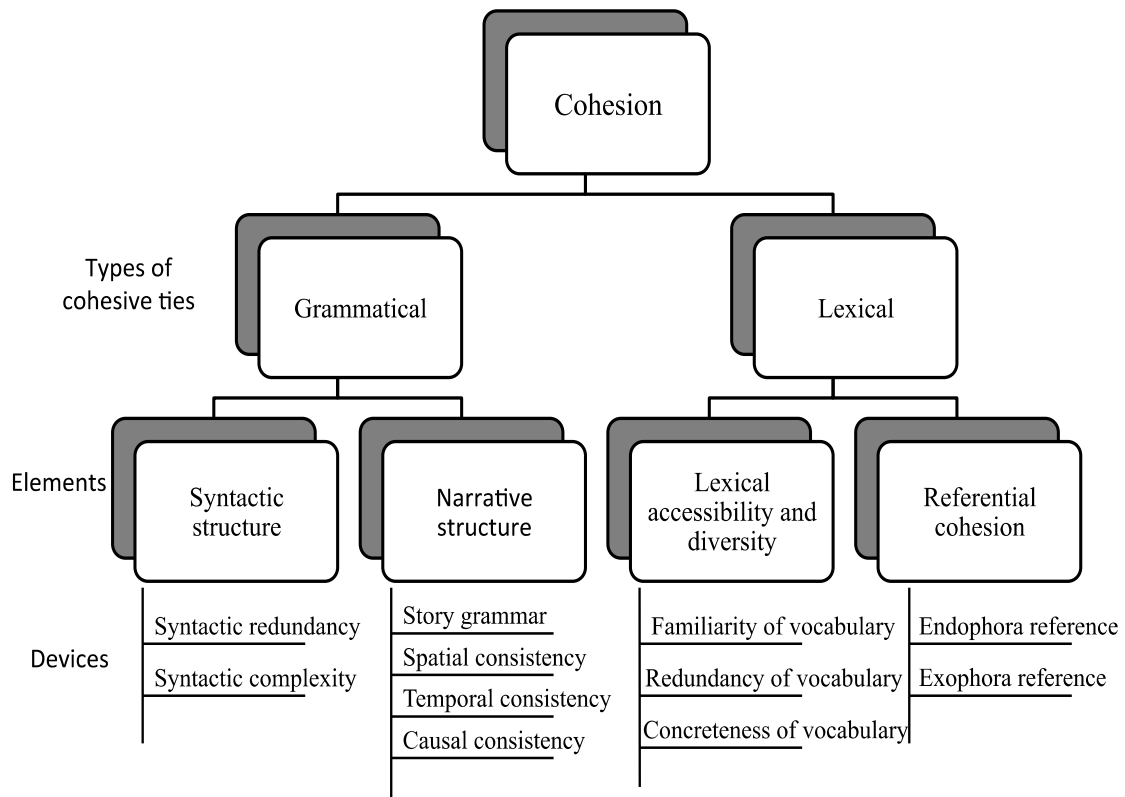


Figure 2. Integrated model of text cohesion.

(Graesser, McNamara, & Kulikowich, 2011). These two components – syntactic redundancy and syntactic complexity – capture the effect of syntactic structure on the cohesion of the text as a whole.

Syntactic redundancy is the repetition of sentence structures across the text (Stanovich, 1980). Syntactic redundancy is related to syntactic priming and syntactic parallelism, because the repetition of syntactic structure primes the reader for syntactic processing. This device contributes to the cohesion of a text because it structurally links a series of sentences and allows for efficient processing of text meaning. Additionally, syntactic redundancy contributes to faster processing times (as measured by reading rates), separate from the effects of lexical word repetition (Ledoux, Traxler, & Swaab,

2007). For an example of syntactic redundancy, consider the following sentences: “Ben and Alice had a picnic. Ben and Alice were happy.” While these sentences share lexical features (e.g., the words “Ben” and “Alice”), they also share a grammatical structure – they are both single independent clauses following a subject-subject-verb format. The repeated use of the same sentence structure builds familiarity so that the reader can attend to passage meaning. At a broader level, consider this section: each cohesive device is described in a single paragraph in four steps: a) a description of the device, b) an explanation of the device’s connection to text cohesion, c) an example of the device, and d) means to measure the device. Because this structure is repeated across all paragraphs within the section, the syntactic structure is redundant. Syntactic redundancy can be challenging to measure, but the authors of a program called Coh-Metrix have developed a method of parsing sentences into part-of-speech categories to create a tree-style representation of the syntactic structure, which can be compared to other sentences to obtain a measure of syntactic similarity (Graesser & McNamara, 2011).

A second feature of the syntactic structure of a text, syntactic complexity, refers to the complexity of sentence structures within the text (Graesser, McNamara, Louwse, & Cai, 2004). For example, a syntactically simple sentence may contain just one independent clause, while a syntactically complex sentence may contain multiple independent and dependent clauses. Syntactic complexity contributes to cohesion because the complexity of grammatical structures can support or hinder reader understanding of connections across the text (Pearson, 1974). For example, consider the sentences: “All African elephants have tusks. Their tusks are highly sought by ivory hunters.” Both sentences follow the basic format subject+verb+object. Now, consider

these sentences: “Unlike their Asian counterparts, of which only males grow ivory tusks, all African elephants male and female have tusks. These tusks are popular among artists and have become a target of ivory hunters, who have impacted elephant populations by aggressively hunting tusked elephants.” These sentences utilize more complex structures, including the use of multiple clauses in a single sentence. While the second example may be more informative to a skilled reader, the use of complex sentence structures may impact a novice reader’s ability to pick out important information or understand how pairs of sentences are related. Syntactic complexity is generally measured by counting the number of words per sentence; longer sentences are generally indicative of more complex syntactic structures, while shorter sentences generally capture simpler syntactic structures. Some researchers (see Graesser, McNamara, & Kulikowich, 2011; Graesser, McNamara, Louwerse, & Cai, 2004) have also quantified syntactic complexity by measuring the number of causal verbs, intentional actions or events, syntactic similarity, type-token ratio (for each word, the type-token ratio is one divided by the number of occurrences of the word), and mean number of modifiers per noun phrase.

Narrative structure. The narrative structure of a text refers to the consistency of the overall structure of the text (van den Broek & Gustafson, 1999). Narrative structure is included in a discussion of cohesion because it represents the continuity of a recognizable text structure across the entire text. The narrative structure of a text includes four components: story grammar, spatial consistency, temporal consistency, and causal consistency.

Story grammar refers to the use of predictable text components or devices, such as the presentation of characters, a setting, a problem or initiating event, and resolution or

conclusion (Jungjohann, 2008). Story grammar contributes to text cohesion because it allows readers to access schema of how the text should function, and build a mental representation of the text by adding new information from the text to the existing model. Breaks in this global-level consistency may make it more challenging for readers to attend to relevant information and build a clear representation of the text. For example, consider a text that presents the problem resolution without ever stating the initiating event. Such a text may impact a less skilled reader's comprehension of that problem resolution and its role in the overall text. Story grammar may be measured using a qualitative analysis of the text guided by structured questions or checklists.

Spatial consistency refers to the maintenance of orientation in space across the text (Zwaan, Radvansky, Hilliard, & Curiel, 1998; Tapiero, 2007). Spatial consistency can be achieved by maintaining a single spatial orientation or following a logical and explicit progression of space. Spatial consistency contributes to text cohesion because it creates spatial links between sentences and across the text. For example, consider the following sentences: "The detective entered the museum through large, imposing doors. Inside the foyer, the cool drafts and low, rumbling echoes only added to the detective's sense of anxious anticipation." These sentences demonstrate the maintenance of spatial consistency because the character's orientation in space – in the first sentence she enters the building and in the second sentence she is inside the foyer – is logically connected across the sentences. Spatial consistency may be measured by counting the number of spatial indicators – words that provide information about space like "inside" or "over" – in the text.

Temporal consistency refers to the continuity of time across the text (Zwaan et al., 1998; Zwaan, 1996; Tapiero, 2007). Texts that maintain temporal consistency present the passage of time clearly and explicitly, as opposed to jumping from various periods in time or presenting time ambiguously. Temporal consistency contributes to text cohesion because it creates temporal links between sentences and across the text. An example of temporal consistency is highlighted in the following sentences: “Lucille’s day began with a large bowl of oatmeal and a glass of orange juice. After breakfast, Lucille showered and dressed for her big day.” Temporal consistency is maintained in these sentences because events occur sequentially and the passage of time is made explicit to the reader. Temporal consistency may be measured by counting the frequency of temporal connectives like “next,” “before,” or “after.”

Finally, causal consistency refers to the maintenance of a logical cause and effect structure in the text (Zwaan et al., 1998; Tapiero, 2007). Causal consistency contributes to text cohesion because it allows the reader to link initiating ideas or actions with the resulting cause and generate causal inferences while reading (Zwaan & Radvansky, 1998). The following sentence illustrates causal consistency by explicitly stating the causal relationship between idea units: “Elijah went to the grocery store because he was out of milk.” The two clauses – “Elijah went to the grocery store” and “he was out of milk” – are connected by the word “because,” which identifies that Elijah going to the store is the effect of being out of milk. Causal consistency may be measured by counting the frequency of causal connectives – words that indicate causal relations between ideas – such as “because,” “therefore,” and “as a result of.”

Lexical cohesion. Lexical cohesion refers to ties and connections between elements of the text that are due to lexical similarity (Morris & Hirst, 1991). Lexical cohesion preserves the continuity of word meaning across text through the use of lexically similar ideas. Two primary components of lexical cohesion are lexical accessibility and diversity and referential cohesion.

Lexical accessibility and diversity. Lexical accessibility and diversity captures the extent to which the vocabulary in the text is understandable to the reader (Graves & Graves, 2003). This “understandability” of vocabulary is influenced by the familiarity of the vocabulary, redundancy of vocabulary, and the concreteness of vocabulary.

The familiarity of text vocabulary refers to how familiar vocabulary is to the reader (Graesser, McNamara, Louwense, & Cai, 2004). While text familiarity will vary by reader due to background knowledge, there are features of vocabulary familiarity that are inherent to the text; specifically, the commonness or uniqueness of words in discourse. The familiarity of text vocabulary is related to text cohesion because unfamiliar vocabulary may hamper a reader’s ability to connect thoughts and ideas across the text. For example, texts that use challenging, unique, or content-specific vocabulary to describe a common construct, such as “cephalopod” for “squid” may impact a reader’s ability to integrate the vocabulary to other information provided in the text due to a lack of familiarity. Vocabulary familiarity can be evaluated by measuring word frequency in the written language based on a corpus of available text.

The redundancy of vocabulary in the text refers to the repetition of vocabulary across the text. The redundancy of vocabulary, or reiteration, as it is termed by Halliday and Hasan (1976), is related to cohesion because it links texts through a common

referent. An example of lexical redundancy can be seen in the following sentences: “She walked carefully through the old building, ducking out of the way of several cobwebs. She stopped at the largest cobweb, where a dark spider sat in the center waiting patiently for her prize.” In these sentences, the word “cobweb” is reiterated through lexical redundancy. The redundancy of vocabulary in a text can be evaluated by measuring the type-token ratio.

Finally, the concreteness of text vocabulary refers to the level of concreteness or abstractness of words in the text (Graesser et al., 2004). Word concreteness is related to cohesion because it affects the reader’s global representation of the text. Word concreteness can be a product of the word itself – such as abstract constructs like “love” or “freedom” – or can be due to insufficient word meaning information – such as the sentence “We saw her duck.” One way in which word concreteness is measured is polysemy, which captures the number of senses or meanings of a word. Words with greater polysemy are more abstract because the word can mean many different things. Additionally, words can be assigned concreteness scores by human raters, as in the Coh-Metrix program.

Referential cohesion. Referential cohesion captures cohesive ties developed through the continuity of reference (Freebody & Anderson, 1983). Specifically, referential cohesion captures elements in the text that are only interpretable by reference to something else. This type of cohesion is a subtype of lexical cohesion because word meaning is derived from previously provided information. Referential cohesion has many components, which can be grouped together as endophora reference and exophora reference.

Endophora reference refers to word meanings that are derived through reference to previously provided information presented in the text itself (Halliday & Hasan, 1976). Endophora reference is a key component of cohesion because it requires the reader to integrate new information (the reference) to previously provided information in the text (the referent). Examples include personal pronouns like “she,” “they,” and “his,” which only have meaning through reference to other words in the text. Endophora reference can be measured by counting things like anaphor overlap (co-reference between pronouns and referent), argument overlap (shared nouns, pronouns, or noun phrases), content word overlap (re-occurrence or overlap of key content words) and stem overlap (shared morphological elements).

Exophora reference can also be described as situational reference because it describes instances in which word meaning is derived through reference to background information and/or conceptually similar vocabulary available in the lexicon (Halliday & Hasan, 1976). Exophora reference is related to text cohesion because it enables the reader to build a coherent representation of the text meaning through the integration of prior knowledge and background information with the text itself. For example, consider the following sentences: “John went for a run. He likes to exercise.” In these sentences, common situational knowledge links words like “run” and “exercise.” Consequently, a reader can interpret the relationship between the two sentences based on situational or background knowledge about the construct “run.” Exophora reference also captures general (non text-based) reference like “One must be polite to others,” in which the meaning of “one” and “others” is not explicitly stated in the text but generally accepted.

Exophora reference can be measured using Latent Semantic Analysis, a method of statistically capturing the semantic relations between words.

Selection of referential cohesion index for study. While all of these components represent aspects of cohesion, referential cohesion was selected for further evaluation in this study. As previously noted, referential cohesion represents the extent to which words and ideas are related across sentences and the entire passage to create explicit connections for the reader (McNamara, Graesser, Cai, & Kulikowich, 2011). Texts that are high in referential cohesion contain words and ideas that overlap across the text, so that connections between text elements are made explicit to the reader (Hiebert & Pearson, 2010).

Referential cohesion was selected for evaluation because it likely captures a meaningful component of text complexity that is omitted from typical text complexity evaluations. First, referential cohesion likely measures something different than readability ratings. Typically, readability ratings capture decoding difficulty (how difficult the words in the text are to decode), semantic difficulty (how rare the words are in the lexicon), and syntactic difficulty (sentence structure). In theory, syntactic difficulty should be able to capture complexity across sentences; in practice, many readability formulas capture this construct by assessing the mean number of words *within* each sentence. Consequently, it is hypothesized that readability measures do not capture how words and ideas connect across the entire text. Second, referential cohesion appears to play an important role in explaining variability in student reading performance. In analysis by Graesser and colleagues (2011), referential cohesion is second only to narrativity in explaining variance in student reading performance (14.1% compared to

18.5%), and was a stronger predictor of reading performance than syntactic simplicity, word concreteness, causal cohesion, verb cohesion, logical cohesion, and temporal cohesion. This suggests that referential cohesion plays an important role in predicting student reading proficiency. Referential cohesion has also been identified as a particularly strong predictor of reading comprehension (McNamara, Graesser, Cai, & Kulikowich, 2011).

Passage Genre

A second feature of text structure and format that may be incorporated into estimates of text complexity is passage genre. Genre is defined as a category of text characterized by similarities in form, style, or subject matter. While genre is generally described qualitatively – for example, text is characterized as either narrative or informational – this qualitative dimension may have an impact on other types of quantitative features of text complexity. For example, existing research has identified systematic differences in oral reading fluency performance (e.g., Briggs, 2012; Saenz & Fuchs, 2002) and passage-specific comprehension (e.g., Cervetti, Bravo, Hiebert, Pearson, & Jaynes, 2009; Best, Floyd, & McNamara, 2008) on narrative vs. informational text. For the purpose of this research, genre was included as a variable of study as a means of better understanding text cohesion. Specifically, passage genre was included for two reasons: 1) to expand upon existing research on the effects of text cohesion on reading performance, and 2) to explore potential interactions between genre and cohesion. First, much of the existing research evaluating the effects of cohesion on reading fluency and/or reading comprehension has used exclusively narrative or exclusively informational text. Consequently, it is unclear whether findings apply to the

other genre. Secondly, research that has examined cohesion across genres suggests that there may be an interaction between cohesion and genre on reading comprehension. For example, Best and colleagues (Best, Ozura, Floyd, & McNamara, 2006) found that high cohesion texts supported reader comprehension better than low cohesion texts for narrative texts, but that there was no difference in comprehension as a function of cohesion for the informational texts. These results suggest that there is a need to better understand how cohesion functions in both narrative and informational texts to support reader comprehension as well as reading fluency.

Purpose of This Research and Hypotheses

The purpose of this study is to evaluate referential cohesion and passage genre as features of text complexity that may enhance the utility and precision of formative assessment tools. Specifically, this design evaluated the effects of referential cohesion and genre on reading rate, accuracy, and passage-specific comprehension on passages deemed equivalent by existing means of quantifying text complexity (i.e., readability formulas). The study design allowed for examination of main effects of referential cohesion and genre as well as interaction effects.

The primary hypothesis was that that readers perform better – read more correct words in one minute, with a higher degree of accuracy, and with better comprehension – on passages with high referential cohesion compared to passages with low referential cohesion, when readability is held constant. Similarly, it is hypothesized that genre acts directly on oral reading fluency, accuracy, and passage-specific comprehension, with increases in all three dependent variables for narrative texts compared to informational texts. It is also hypothesized that a relation exists between the independent variables,

which is the reason for the inclusion of genre in the study design. However, analysis of this hypothesis is beyond the scope of this study.

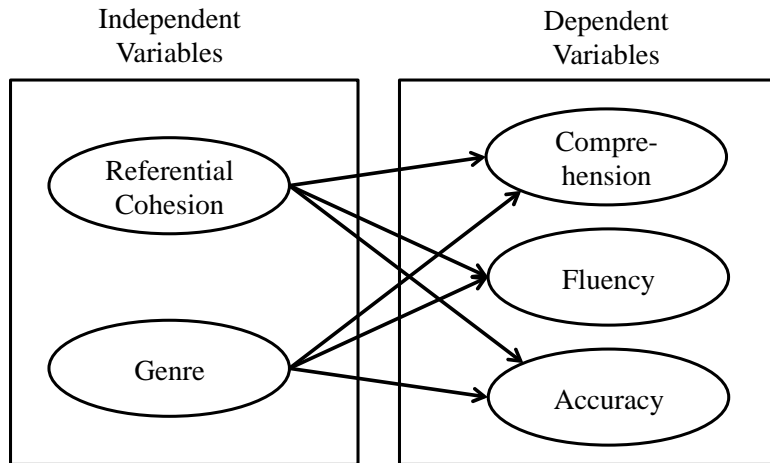


Figure 3. A model of relations between independent variables (genre, referential cohesion) and dependent variables (oral reading fluency rate, oral reading fluency accuracy, passage-specific reading comprehension).

This study design allowed for examination of direct effects between independent and dependent variables.

Direct effects on oral reading fluency rate. It was hypothesized that referential cohesion acts directly on oral reading fluency rate, because high referential cohesion increases the predictability of text, which may lead to increases in word reading speed (e.g., semantic priming). It was also hypothesized that genre acts directly on rate, as narrative texts follow predictable structures that may lead to more efficient reading.

Direct effects on oral reading fluency accuracy. As with rate, it was hypothesized that referential cohesion acts directly on oral reading fluency accuracy by increasing the predictability of text through lexical redundancies. Assuming that students

read words correctly the first time, it was predicted that the repetition of previously mastered words would increase accuracy of reading. It was also hypothesized that genre acts directly on accuracy, as informational texts may contain a greater number of content-specific words.

Direct effects on passage-specific comprehension. It was hypothesized that referential cohesion acts directly on passage-specific comprehension by making explicit the relations between ideas in the text. It was hypothesized that increased explicitness leads to increased passage-specific reading comprehension. It was also hypothesized that genre acts directly on passage-specific comprehension, as narrative texts may be more predictable in structure and consequently lead to greater understanding.

Research questions. Evaluation of these hypotheses was guided by the following research questions:

1. When readability is held constant, do students read more correct words per minute on passages with higher referential cohesion than passages with lower referential cohesion?
2. When readability is held constant, do students read passages with higher referential cohesion with greater accuracy than passages with lower referential cohesion?
3. When readability is held constant, do students perform better on a measure of passage-specific reading comprehension for passages with higher referential cohesion than passages with lower referential cohesion?

Additionally, it was hypothesized that the effects of cohesion may interact with the genre of the passage. This hypothesis included evaluation of main effects of genre

and interactions between genre and referential cohesion, as described in the following questions:

4. When readability and referential cohesion are held constant, do students read more correct words per minute on narrative texts than informational?
5. When readability and referential cohesion are held constant, do students read narrative texts with greater accuracy than informational texts?
6. When readability and referential cohesion are held constant, do students perform better on a measure of passage-specific reading comprehension on narrative texts than informational?
7. If differences in oral reading performance are noted on high and low cohesion passages (questions 1, 2, and 3), do the effects depend on whether the text is narrative or informational?

CHAPTER II

LITERATURE REVIEW

What Makes Text Difficult?

Before evaluating methods of assessing the difficulty of a text, one must first understand what variables contribute to text complexity. Researchers seem to agree that both text and reader variables not only independently contribute to the complexity of a given text, but also interact to affect text complexity (Anderson & Pearson, 1984; Common Core Standards Initiative, 2010; McKeown, Beck, Sinatra, & Loxterman, 1992; Hiebert & Fisher, 2007). In a synthesis of existing work on text complexity, Graves and Graves (2003) summarize text complexity factors as including ten features: vocabulary, sentence structure, passage length, elaboration, coherence and unity, text structure, familiarity of content and background knowledge required, audience appropriateness, quality and verve of the writing, and interestingness. Graves and Graves divide these factors into two broad categories – text-based and reader-based features.

Text-Based Features of Text Complexity

Graves and Graves (2003) identify vocabulary, sentence structure, length, elaboration, coherence and unity, and text structure as text-based features of text complexity. The remaining features are described as reader-based features of text complexity, and consequently will not be the focus of this review. These six text-based features can be grouped into three domains: semantic complexity (vocabulary, elaboration), syntactic complexity (sentence structure, length), and coherence and cohesion (coherence and unity, text structure). An additional domain, decoding difficulty, is added to this discussion, as Graves and Graves fail to capture this feature of

complexity in their domains. Text-based variables are identified as features inherent to the text itself. While Graves and Graves acknowledge that no text is completely independent from the reader, these features largely describe variability captured in the text itself, detached from reader skills and knowledge.

Decoding difficulty. At its most basic level, the complexity of a passage is affected by how difficult the words in that passage are to decode. Word recognition is a foundational reading skill (Archer, Gleason, & Vachon, 2003), and while decoding skills and deficits vary from reader to reader, there are features of the word itself that can support or hinder efficient decoding (Hiebert, 1998). For example, English has a fairly opaque or deep orthography, in which grapheme-phoneme correspondences are not always consistent (Baker, Stoolmiller, Good, & Baker, 2011; Ehri, 2005). Multiple graphemes may represent the same sound, as in hay, late, and sleigh (all are pronounced with the long ā sound, but spellings vary). Conversely, one grapheme may represent multiple phonemes, as in the letter “a” in ago, apple, and wary. Additionally, longer words may be more difficult for readers to decode (Powell-Smith, Good, & Atkins, 2010).

Semantic difficulty. While decoding difficulty refers to how difficult words in the passage are to decode, semantic difficulty refers to how difficult words are to understand. Research suggests that semantic difficulty is a strong predictor of overall passage difficulty (Graves & Graves, 2003). In general, texts containing lots of challenging words tend to be more challenging overall. However, the semantic difficulty of a passage is not just about how “easy” or “hard” individual words are; rather, it is the appropriateness of the words for the context that impacts a reader’s ability to comprehend

a passage. Consequently, texts with “harder” vocabulary may be easier for a reader to comprehend if that vocabulary is necessary to convey the author’s meaning.

Syntactic difficulty. Syntactic difficulty captures the structural features of the text, both at the sentence-level and at the passage-level. At the sentence-level, syntactic difficulty refers to sentence length and complexity. While shorter sentences are generally considered to reduce the difficulty of a text (see description of readability formulas below), short sentences that fail to resemble spoken language and lack connectives between ideas may be more difficult for readers to comprehend. There is also evidence that texts with varied sentence structure may be easier to read and comprehend than those with limited, short sentence structures (Hiebert, 1998). At the passage-level, syntactic difficulty captures the organization of the text as a whole. This includes how ideas are sequenced, the use of illustrations, headings, and the expression of relationships between ideas (Risko & Walker-Dalhouse, 2011; Beers & Nagy, 2009). Genre also contributes to passage-level text structure, as narrative and expository texts tend to be organized differently (Graves & Graves, 2003). Finally, Graves and Graves argue that overall passage length contributes to difficulty, as length may prompt expectations for the reader, and may be indicative of text structure.

Coherence and cohesion. Coherence serves as a bridge between text-based and reader-based features of text complexity, as it captures how the text supports the reader’s formation of a mental representation of the text. Coherence includes many features of text complexity described above, such as text organization, sequencing, explicitness of relationships, and language (McKeown, Beck, Sinatra, & Loxterman, 1992). Coherence also captures the extent to which the reader must make inferences in order to understand

the meaning of the text. The coherence of a text cannot be measured, because coherence by definition is influenced by reader skills and background knowledge. However, the text-based features that support coherence are described as text cohesion factors and can be measured (Graesser, McNamara, Louwerse, & Cai, 2004). Text cohesion includes the kinds of variables that influence reader understanding of the text, such as co-reference and overlap, the incidence of connective words, connectives between causes and effects, and semantic similarity of words in a passage (McNamara, Louwerse, McCarthy, & Graesser, 2010). Highly cohesive texts support reader comprehension by increasing comprehensibility, while low cohesion texts require the reader to use his/her skills and background knowledge to piece together the meaning of the text.

Reader-Based Features of Text Complexity

Graves and Graves (2003) group the remaining four variables – familiarity of content and background knowledge required, audience appropriateness, quality and verve of the writing, and interestingness – together as reader variables, because they involve the reader and the reader’s interaction with a text. In general, these features capture what the reader brings to the text: background knowledge about the nature of reading and the content of the text, age and developmental level, and interests and preferences. Because these features vary from reader to reader and can’t necessarily be captured or controlled, reader-based contributions are generally omitted from text complexity measurement approaches.

Approaches to Evaluating Text Complexity

Researchers and practitioners have developed frameworks for evaluating text-based features of text complexity, including readability formulas, the National

Assessment of Education Progress (2008) framework, and the Common Core Standards (2010) framework. These approaches evaluate text complexity in slightly different ways, but all identify means of quantifying text-based features of complexity.

Readability formulas. One approach to assessing the difficulty of a text is to focus strictly on the quantitative features of the text. In this approach, the language elements present in a selection of text are counted and used to predict reader performance on a criterion measure of comprehension. These scores can then be used to create readability formulas, which can be applied to new texts to determine the text complexity.

Readability formulas typically assess text complexity on three of the four domains described above: decoding difficulty, semantic difficulty, and syntactic difficulty (Powell-Smith, Good, & Atkins, 2010). Decoding difficulty is generally assessed by counting the number of letters or syllables in each word. Semantic difficulty is generally assessed by counting the number of low-frequency or rare words in the passage – as determined by a list of high-frequency words (e.g., the Dale list) or a corpus of text (e.g., Lexile Framework for Reading). Syntactic difficulty is generally assessed by counting sentence length – either the number of words per sentence, or the number of syllables per sentence. Together, these scores are combined to create an overall indicator of the complexity of the passage.

While readability formulas have a long history in the assessment of text complexity, they have not gone without criticism. First, readability formulas are designed to allow for predictions of student comprehension skills; however, in practice readability scores don't relate strongly to student comprehension of the text (Hiebert, 2011). For example, passages with short sentences and frequent words would lead to

“easier” designations of text complexity, but such texts may not actually be easier to comprehend (Hiebert, 2011). Beck and colleagues (Beck, McKeown, Omanson, & Pople, 1984) illustrate this argument in a study focusing on how text revisions impact student comprehension of the passage. In this work, researchers manipulated stories from basal readers to be more coherent from both bottom-up (e.g., altering specific wording or phrasing) and top-down (e.g., re-organizing events to be more conceptually consistent) perspectives. In doing so, researchers increased both the number of words in the stories as well as the readability ratings, which increased by one grade level for each story. Participants in the Beck et al. study were then presented with either the original or the revised passages, and were directed to read the passages as they would during a basal reading lesson. After completing each passage, comprehension was assessed using a measure of passage recall and a multiple-choice comprehension test. Beck and colleagues found that students who read the revised passages, which had higher readability estimates (i.e., were less readable) than the original passages, scored higher on both the recall and multiple-choice comprehension assessments. Specifically, students that read the revised passages recalled more information that was central to the passage narrative, and answered more comprehension questions than students in the control group. These findings suggest that something other than text readability – namely, text coherence and cohesion – affects student comprehension of text. Given that comprehension is the goal of reading, these findings cast doubt on the ability of readability formulas to fully capture the text elements that contribute to text complexity.

More recently, McNamara (2001) examined the relationship between coherence and reader skill level. While this research focused on how coherence interacts with

reader skills to support comprehension, which is beyond the scope of this discussion, it also sheds some light on the relationship between coherence and readability. Like Beck, McNamara manipulated the cohesiveness of science texts to be less or more cohesive. Passages were generally low in cohesiveness, so the majority of revisions sought to increase text coherence by: replacing pronouns with nouns, adding elaborations, inserting words to connect relationships between ideas (e.g., however, because), increasing content overlap across sentences, inserting headings, adding explicit topic sentences, and rearranging sentence order. McNamara found meaningful differences in readability between the high- and low-coherence texts – while the high-coherence texts contained 900 words in 50 sentences, the low-coherence texts contained 650 words in 48 sentences. As a result, readability grade-level estimates ranged from 11.2 (high-coherence) to 9.3 (low-coherence). These differences in readability suggest that the low-coherence passages should be easier to read – in other words, to comprehend – than the high-coherence passages. However, McNamara found that low-coherence passages were only easier to read if readers had high levels of background knowledge about the topic. Without such pre-existing knowledge, readers benefitted from reading text with high levels of cohesion, even if readability was more difficult as a result.

Second, there is evidence to suggest that the variables included in readability formulas may contribute to reading performance differently based on the type of text. Research by Cohen and Steinberg (1983) has examined the semantic difficulty indicator used in readability formulas within the context of science textbooks. Traditionally, readability formulas have used word lists like the Dale List of 3000 Familiar Words (Dale & Chall, 1948) to identify rare or unfamiliar words. However, elementary science

textbooks tend to use words that do not appear on such word lists but appear repeatedly and are defined within the context of the text, suggesting that these words are not truly unfamiliar. Consequently, readability formulas using this approach to capture semantic difficulty may overestimate the text complexity of science textbooks. Cohen and Steinberg evaluated this argument by analyzing the types of unfamiliar words present in elementary science textbooks. Using three commercially available elementary science textbooks, the researchers categorized unfamiliar words (which, according to the Dale List, made up almost 15% of evaluated words) into three categories – technical (words that were the subject of the text or were defined in the text), technical support (words that are not as recognizable as technical words but are commonly used in science), and non-technical (words that are not common in science or central to the content of the text). Cohen and Steinberg found that the majority of unfamiliar words were technical words, and that the inclusion of these technical words in the percent of rare/unfamiliar words count in many readability formulas inflated readability estimates for science texts.

Similarly, readability estimates may fail to capture the unique contributors to text complexity that occur in other specialized texts, like poetry and early reading texts. According to Foorman (2009), meaning in poetry texts is often tied to language and text structure, rather than word frequency or sentence length – two features central to readability estimates. Foorman argues that poetry may include vocabulary that would be considered unfamiliar based on word lists or banks, but readers can extract meaning from the text by relying on text structure. Thus, the difficulty of such a passage may be inaccurately captured by readability estimates, which fail to assess text beyond surface-level characteristics. In contrast, early reading texts tend to contain a number of high-

frequency words, which would correspond with lower readability estimates; however, evidence suggests that the majority of words included in early reading texts only occur once, and consequently fail to provide enough exposures for such high-frequency words to be integrated into student sight word vocabularies (Foorman, Francis, Davidson, Harm, & Griffin, 2004, as cited in Foorman, 2009). As a result, even texts with a low percentage of rare/unfamiliar words can be challenging for early readers.

Finally, some researchers have questioned the validity of a single score in capturing the complexity of a text, particularly longer texts. For example, the Lexile Map rates the narrative text *Pride and Prejudice* as a 1100 Lexile, corresponding with approximately an 8th-12th grade level. However, individual chapters of the text show great variability in readability estimates, from 670 (3rd grade) to 1310 (college) (Hiebert, 2011). This suggests that a single readability estimate cannot capture text complexity across a variable text. Additionally, the use of a single measure of text complexity limits the treatment utility of using readability estimates to select appropriate texts (Graesser, McNamara, & Kulikowich, 2011). While a placement system like the Lexile Framework for Reading may place two readers at the same level, their individual needs may not be met by the same text.

2009 NAEP reading framework. The National Assessment of Education Progress (NAEP), or “Nation’s Report Card,” is an ongoing effort to collect data on national student achievement in academic subject areas such as reading and mathematics (National Assessment Governing Board, 2008). The NAEP is administered to a demographically representative sample of students in grades 4, 8, and 12, and can be used to assess student achievement at the national and state levels as a whole and for targeted

subgroups. The most recent NAEP reading assessment was administered in 2009 in accordance with the 2009 NAEP Reading Framework.

A central assumption of the NAEP is that text increases in complexity from grade 4 to grade 12. Consequently, an evaluation of passage complexity is critical to the selection of appropriate testing materials. According to the Framework, selected texts must be “of the highest quality, evidencing characteristics of good writing, coherence, and appropriateness for each grade level” (p. 27) and must become “successively more complex” (p. 16) at each grade level. In general, the complexity of potential passages is evaluated by considering the following variables: passage length, quality of writing, interestingness, writing style, text organization, sentence structure, vocabulary, supplementary materials (e.g., definitions of technical terms), and elaboration. Specific text structures and features are presented for each grade level and each type of text included in the assessment (fiction, literary nonfiction, poetry, exposition, argumentation/persuasive text, and procedural text/documents). Evaluation of the identified contributors to text complexity is based primarily on expert judgment, but must also include story and concept mapping and at least two research-based readability formulas.

Common Core Standards framework. Text complexity is a central component in the Common Core Standards (2010), an initiative towards universal standards in English/Language Arts, History/Social Studies, Science, Math, and Technical Subjects. These standards are based on existing educational research, and are designed to support schools in targeting the skills students need for college and workplace success. Embedded within the Common Core Standards is the expectation that students read and

comprehend text of increasing complexity as they progress through their schooling. In order to assess the complexity of a given text, the Common Core Standards describe a three-fold evaluation approach. Within this model, each individual evaluation contributes to the overall evaluation of the text.

First, the Common Core Standards recommend a quantitative evaluation of text complexity. The Common Core Standards do not endorse any particular method of quantitative analysis; rather, they suggest a thoughtful review of existing tools to best match the measurement tool with the purpose. Some of the quantitative tools suggested for review include traditional readability formulas, newer readability methodologies like the Lexile Framework, and the Coh-Metrix system for assessing text cohesion. The Common Core Standards caution users that many quantitative measures may underestimate complex text (e.g., text with complex ideas, multiple meanings, etc.), and consequently it is important to remember that quantitative analysis is just one component of a thorough evaluation of text complexity.

Second, the Common Core Standards recommend a qualitative evaluation of the text. This evaluation includes analysis of four factors: levels of meaning, text structure, language conventionality and clarity, and knowledge demands (see Table 1 for additional information about each factor). These factors are not intended to be captured quantitatively; rather, the Common Core Standards recommend using evaluator judgment and expertise to determine the contributions of each of these factors to the overall difficulty of the passage. The Common Core Standards stress that quantitative measures alone do not capture all elements of text complexity, and this qualitative evaluation is a necessary supplement to quantitative analysis.

Table 1.

Qualitative Dimensions of Text Complexity Included in the Common Core Standards Framework.

Dimension	Less Complex	More Complex
Levels of meaning or purpose	Single layer of meaning Explicitly stated purpose	Multiple levels of meaning Implicit purpose, may be hidden or obscure
Structure	Simple Explicit Conventional Events related in chronological order Traits of a common genre or subgenre Simple graphics Graphics unnecessary or merely supplementary to understanding the text	Complex Implicit Unconventional Events related out of chronological order Traits specific to a particular discipline Sophisticated graphics Graphics essential to understanding the text and may provide information not otherwise conveyed in the text
Language Conventinality and Clarity	Literal Clear Contemporary, familiar Conversational	Figurative or ironic Ambiguous or purposefully misleading Archaic or otherwise unfamiliar General academic and domain-specific
Knowledge Demands: Life Experiences	Simple theme Single theme Common, everyday experiences or clearly fantastical situations Single perspective Perspective(s) like one's own	Complex or sophisticated themes Multiple themes Experiences distinctly different from one's own Multiple perspectives Perspective(s) unlike or in opposition to one's own
Knowledge Demands: Cultural/Literary Knowledge	Everyday knowledge and familiarity with genre conventions required Low intertextuality (few if any references/allusions to other texts)	Cultural and literary knowledge useful High intertextuality (many references/allusions to other texts)
Knowledge Demands: Content/Discipline Knowledge	Everyday knowledge and familiarity with genre conventions required Low intertextuality (few if any references to/citations of other texts)	Extensive, perhaps specialized discipline-specific content knowledge required High intertextuality (many references to/citations of other texts)

Third, the Common Core Standards propose an evaluation of reader and task considerations. While the previous two evaluations focus on text-based variability in text complexity, this evaluation shifts focus to reader variables such as cognitive skills, motivation, knowledge, and experiences (RAND Reading Study Group, 2002). Additionally, this evaluation should include a review of the complexity of the academic task assigned, as the academic expectation (e.g., skimming vs. studying) may impact how challenging the text is for that particular purpose. For example, a science text with organizing headings and highlighted vocabulary may be easier for the purpose of skimming for key points and more challenging for the purpose of identifying specific information.

These three components are then combined to assign a grade band to the text. Unlike readability indices, which assign a quantitative readability score to the passage, the Standards provide recommended placement in one of the following grade bands: 2-3, 4-5, 6-8, 9-10, and 11-college/career readiness level. Because these bands span multiple grades, it is expected that students in the lower range require scaffolding and support to comprehend the text, while students at the upper range should be able to read and comprehend the text independently.

While the Common Core Standards specify that all three components are equally important in an evaluation of a text's complexity, each of the three methods should not be given equal weight for every text. Professional judgment is required to determine how appropriate each assessment is for the selected text. For example, the authors argue that a quantitative tool such as a readability formula may provide valuable information in evaluating a dramatic text, but may fail to capture the difficulty of a poem. A thoughtful

evaluation of text complexity should include consideration of how to weigh each component based on the individual text.

Text Cohesion: A Potential Contribution to the Evaluation of Text Complexity

Text cohesion captures the extent to which a text hangs together as a coherent whole (Morris & Hirst, 1991). Cohesion is different from the construct of coherence, which describes the mental picture a reader constructs during reading based on both the text and background knowledge. While coherence addresses the interaction between the text and the reader in constructing meaning, text cohesion focuses exclusively on the supportiveness of the text in facilitating comprehension. Consequently, the cohesive features of a text can be evaluated, and may support our understanding of the complexity of the text structure.

Effects of cohesion as a whole. Support for the role of cohesion in oral reading performance is provided by comparisons of student performance of tasks of word list reading fluency and passage reading fluency. This allows researchers to evaluate the effect of text cohesion – inherent in the passage – versus the absence of cohesion – inherent in the word lists. If oral reading fluency is strictly a product of efficient decoding skills and cohesiveness of text plays no role, performance should be similar on a passage or the same words in a list.

Jenkins and colleagues (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003a; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003b) contrasted word-list and passage reading performance for students across a range of reading skills. Students were administered two brief, fluency-based measures – one of word-list reading skill and another of passage reading skill – and a group administered

test of reading comprehension. Mean fluency score for passage reading was significantly higher than the mean fluency score for word list reading. Furthermore, regression analyses indicated that passage fluency uniquely explained 42% of the variance in reading comprehension scores, while word-list fluency uniquely explained only 1%. These findings suggest that the cohesiveness inherent in the passage may contribute to comprehensibility. Researchers performed additional analysis using passage fluency as the outcome variable and reading comprehension and word-list fluency as predictors. They found that word-list fluency uniquely explained 11% of the variance in passage reading while reading comprehension uniquely explained an additional 28% above and beyond word-list fluency. This may be explained by the fact that word lists by nature lack cohesion – words are unrelated and unconnected. Passages, on the other hand, contain more cohesion, which may be contributing to student ability to read connected text with appropriate rate and accuracy. These results indicate that cohesive text both 1) facilitated oral reading fluency, and 2) increased the relation between fluency and comprehension.

More recent work using a sample of students receiving both English- and Spanish-language instruction evaluated the contribution of comprehension in explaining passage fluency within and across languages (Baker, Stoolmiller, Good, & Baker, 2011). In this work, participants were assessed using measures of word reading fluency, passage reading fluency, and comprehension in order to evaluate relations between skills. In addition to evaluating performance on all measures across languages, researchers were interested in examining the effects of comprehension on passage reading fluency performance, after controlling for word reading skills. Results suggest that passage

meaning and context contribute to oral reading fluency. First, researchers found that scores on the measures of word reading fluency were significantly lower than scores on passage reading fluency in both languages, indicating that a cohesive passage context contributes to oral reading rate. Second, correlations between passage reading fluency and comprehension scores were significantly higher than correlations between word reading fluency and comprehension in both languages. These findings indicate that factors like cohesion both increase the comprehensibility of text and are likely to increase oral reading fluency.

Effects of referential cohesion. Evidence suggests that, in addition to global cohesion, specific cohesive elements contribute to text complexity. Specifically, measures of referential cohesion have been linked to differences in reading performance. In one study, researchers quantified features of text cohesion to create two cohesion indices: referential overlap (referential cohesion) and vocabulary accessibility (Duran, Bellissens, Taylor, & McNamara, 2007). The referential overlap score captured the degree to which a text displayed conceptual redundancy, or relatedness between sentences. Vocabulary accessibility went beyond typical measures of word frequency to capture word familiarity, ambiguity, and abstractedness. These indices were selected because they were hypothesized by the author to be key features of text complexity. While both scores were significantly correlated with a measure of readability, the correlations were low to moderate (.32 to .54) suggesting that the cohesion indices were measuring a similar but not identical construct as the readability estimates. Scores on these cohesion indices were then used to group four texts on two topics as easy or difficult. Participants read all four passages, and measures of reading rate and passage

retell were obtained. Results indicate significant differences in both reading times and retells for easy vs. difficult texts, suggesting that these measures of text cohesion are capable of distinguishing between high and low complexity texts.

Additionally, work by Posner and Snyder (1975) supports a relation between comprehension and fluency by asserting that the context of a word – which refers to the relations between words in the text, a key feature of text cohesion – facilitates increased word recognition through the activation of semantic networks. While Posner and Snyder did not describe it as such, this facilitation of word reading by the context can be described as a type of exophora reference, because situational knowledge activates networks of similar words (e.g., as in Latent Semantic Analysis). According to Posner and Snyder, each word processed by the reader activates a network of semantically related words, and thus speeds recognition of any subsequent stimuli that fall within the network. As the reader continues to read, the conscious expectancy process inhibits the retrieval of unexpected words. In theory, these processes should support more efficient reading of words that carry similar meaning as opposed to those read out of context or meaning. Referential cohesion is tied to this process because cohesion captures linkages between words through endophora and exophora reference. When texts are highly cohesive, readers can anticipate what is coming because the entire text is constructed as a unified whole, while texts that lack cohesion may disrupt the expectancy process by lacking clear relations between words and ideas.

Stanovich and West (1981) found support for the Posner-Snyder expectancy theory in their evaluation of sentence context on word recognition. Sentence context is one component of cohesion, as context facilitates relations between words and ideas. In

this work, Stanovich and West manipulated the decoding difficulty of words in sentences and measured reader reaction times. Results demonstrated an interaction between word difficulty and cohesiveness of text – the more difficult the words, the greater the effect of text cohesiveness on reaction times. When context sentences were more cohesive, readers read faster, while context sentences with less cohesion increased reading times. Stanovich and West argue that reading speed increased in cohesive contexts because semantic activation occurs while readers decode difficult words. Consequently, the difficult words act to prime the reader to remaining words in the sentence, a process which is only effective if words in the text are related to the cohesive whole.

In addition to these studies on the effects of referential cohesion at a global level, a number of studies explore the role of multiple referential cohesion devices in comprehension. In these types of studies passages are re-written to improve cohesion as well as other features of text complexity (such as syntactic structure), and student performance is compared on original and revised texts. For example, one study by McNamara and colleagues (1996) evaluated the effects of cohesion on reading performance by making the following revisions: replacing ambiguous pronouns with nouns or noun phrases (a component of endophora reference, a referential cohesion device), connecting unfamiliar concepts to familiar ones through elaboration, adding connectives between sentences, increasing argument overlap (a type of endophora reference), and manipulating the syntactic structure of the text by adding topic headers and topic sentences. While this type of research makes it challenging to isolate the specific effects of any one element of cohesion or cohesive device, these studies do support the inclusion of these elements and devices in the integrated model of cohesion

(see Table 2 for a summary of selected research on multiple cohesive devices). Studies have identified many of the cohesive devices included in the integrated model to be related to improvements in reading comprehension when studied in combination with other devices. Consequently, these elements of cohesion and cohesive devices have been identified as meaningful components of text cohesion in the integrated model of cohesion.

Cohesion and Readability: Related but Distinct Constructs

Research suggests that readability formulas and measures of text cohesion do not evaluate text in the same way. In a recent study, Hiebert (2011) used the Lexile Framework Lexile score and component scores (sentence length and word frequency) and a measure of referential cohesion derived from the Coh-Metrix framework (Graesser, McNamara, Louwerse, & Cai, 2004) to evaluate exemplar texts as identified by the Common Core Standards (2010). Her findings indicate that rank orderings of text complexity differ fairly dramatically depending on the metric used. For example, a text identified as the “easiest” or least complex by the overall Lexile score was ranked the “hardest” or most complex text by the referential cohesion score. Additionally, correlations between referential cohesion and the Lexile measures were not statistically significant, implying that referential cohesion is capturing something different than the Lexile readability measures. Consequently, there is a need to further evaluate the contributions of text cohesion to text complexity.

Interactions Between Cohesion and Genre

While cohesion is the primary variable of interest in this study, genre was also selected as a variable for this research. Genre was included in the design because there is

Table 2.

Results of Selected Studies Evaluating the Effects of Revisions to Improve Referential Cohesion on Reading Comprehension Performance.

Study	Devices evaluated	Findings
Ozuru, Briner, Best, & McNamara, 2010	Consistency, endophora reference (anaphor reference, argument overlap)	Text revisions were related to higher quality responses when asked to self-explain the text. However, performance on open-ended comprehension questions was higher for original (low cohesion) texts.
Ozuru, Dempsey, & McNamara, 2009	Consistency, endophora reference (anaphor reference), exophora reference (content word overlap) , semantic and syntactic structures	Revised texts were associated with improved comprehension on passage-specific open-ended comprehension questions. Interactions were found between reader skill level and cohesion on comprehension.
McNamara, 2001	Consistency, endophora reference (anaphor reference), exophora reference (content word overlap) , semantic and syntactic structures	Revised texts were associated with improvements in passage-specific comprehension questions for students with low background knowledge.
Vidal-Abarca, Martinez, & Gilabert, 2000	Causal consistency, endophora reference (argument overlap)	Revisions to argument overlap alone did not improve comprehension as measured by inference questioning and recall. However, revisions to both devices resulted in larger improvements in comprehension than revision to causal connectives alone.
McNamara et al., 1996	Consistency, endophora reference (anaphor reference, argument overlap) , semantic and syntactic structures	Revised texts were associated with improvements in comprehension as measured by a recall task, open-ended questions, and a card sorting task. Interactions were noted between background knowledge and text cohesion.
Britton & Gulgoz, 1991	Endophora reference (anaphor reference, argument overlap) , syntactic structure	Revised texts were associated with improvements in comprehension as measured by a recall task, multiple choice questions, and a keyword association task.
Beck, McKeown, Omanson, & Pople, 1984	Background knowledge, conjunctions, content problems, endophora reference (anaphor reference,) semantic and syntactic structures	Revised texts were associated with improvements in comprehension as measured by a recall task.

Note: Devices that are highlighted in bold are identified as components of referential cohesion.

evidence to suggest that cohesion may impact student reading performance differently based on the genre of the passage. For example, Best and colleagues (Best, Ozura, Floyd, & McNamara, 2006) had students read two narrative and two expository texts selected from school textbooks. All passages were re-written to include both a high cohesion and a low cohesion version (which included manipulations of referential cohesion); students read one high cohesion text within each genre and one low cohesion text within each genre. Comprehension was then measured using a multiple choice question format. Results indicated a main effect for genre, with students earning higher comprehension scores on narrative texts than on expository texts. Researchers also found a main effect for cohesion, with students demonstrating greater comprehension on high cohesion passages than low cohesion passages. Finally, researcher found a significant interaction between genre and cohesion. Students demonstrated greater comprehension on high cohesion narrative passages than low cohesion narrative passages, but did not perform differently on high cohesion versus low cohesion expository texts. These results suggest that cohesion supports reader comprehension for narrative texts, but may be less important for expository texts. Further study is necessary to better understand the relationship between genre and cohesion on comprehension as well as oral reading fluency, and implications for formative assessment.

Quantifying Text Cohesion Using Coh-Metrix

Coh-Metrix was developed to assess text beyond the two to three components typically included in readability analysis; Coh-Metrix provides quantitative information on 54 domains of text cohesion and readability, including lexicons, syntax, and latent semantic analysis (LSA) (McNamara, Louwrese, McCarthy, & Graesser, 2010). These

variables are categorized into five broad indices: 1) readability, 2) general word and text information (characteristics of words in the text, such as frequency of usage), 3) syntax (syntactic complexity, syntactic composition, and frequency of the syntactic classes in text), 4) referential and semantic indices (relationships between words in the text), and 5) situation model dimensions (aspects of the text that contribute to a reader's mental model). These indices are designed to analyze text on multiple levels of language and discourse, consistent with multilevel theoretical frameworks of text comprehension (Graesser, McNamara, and Kulikowich, 2011).

Research on the Coh-Metrix tool suggests that the program is capable of differentiating between texts with high cohesion and those with low cohesion, and captures something different than text readability. In one study, the Coh-Metrix authors manipulated natural texts (i.e., texts culled from existing literature such as textbooks and encyclopedias) to create two versions of each passage: one that was highly cohesive, and another that lacked text cohesion (McNamara, Louwrese, McCarthy, & Graesser, 2010). As many features of text cohesion are available in the Coh-Metrix program, researcher selected four indices – LSA, co-reference (referential cohesion), connectives, and ratio of incidence of causal connectives to change-of-state verbs. Results indicate that readability formulas failed to differentiate between high and low cohesion texts while the Coh-Metrix tool successfully differentiated between high cohesion and low cohesion texts on all of the selected indices. These findings support the validity of Coh-Metrix in assessing cohesion and the sensitivity of the tool to discriminate between texts.

Other scholars have suggested that Coh-Metrix is capable of differentiating between texts that other measures of text difficulty might deem equivalent. In one such

argument, Elfenbein (2011) explored previous research in which passage equivalency was key to parsing out specific text complexity effects. Elfenbein described the work of McKoon and Ratcliff (1992), in which three versions of a passage were developed to be equivalent in passage difficulty but variable in the level of inference required of readers. Central to the design is the equivalency of the three version of the passage, as they provide control for the hypothesis that it is the manipulation of level of inference that impacts reading performance. However, Elfenbein inputted each version into the Coh-Metrix tool, and found a number of linguistic differences between the passages. For example, passages varied on the incidence of connectors and proportion of overlapping content words. These results indicate that the Coh-Metrix program may be more capable of capturing distinctions between text than other means of complexity evaluation.

Finally, research on Coh-Metrix suggests that the detailed information provided by the program produces more accurate estimates of text difficulty than surface-level readability characteristics. In work by Crossley and colleagues (Crossley, Greenfield, & McNamara, 2008), researchers compared the validity of three complexity indices derived from Coh-Metrix in predicting reading difficulty for English Language Learners (ELLs). Coh-Metrix variables were selected to provide information on three domains: lexical (word frequency), syntactic (syntactic structure similarity across adjacent sentences), and meaning construction (content word overlap). The lexical and syntactic indices captured much of the same information as readability formulas, while the meaning construction domain went deeper to capture a key component of referential cohesion. When using performance on a cloze task as a criterion, researchers found that all three predictors accounted for 86% of the variance in cloze performance for the ELL sample. This is an

increase over previous work done by the authors, in which surface-level indices accounted for 72% of the variance explained. These findings suggest that the inclusion of a measure of referential cohesion may allow us to make better predictions about text complexity and student performance.

Summary

The body of evidence suggests that text cohesion contributes to reader performance on measures of fluency and comprehension, and consequently may be an important component of text complexity. Research indicates that readers read more quickly and comprehend better when sentences and passages are more cohesive, and when words are provided in a cohesive context that imparts a meaning or purpose for reading. In short, cohesion matters. Therefore, methods of assessing the complexity of text that focus solely on the decodability, semantic, and syntactic features of the text and not how the words form a coherent whole may not capture the potential impact that cohesion has on reading proficiency. Measures that capture the contributions of cohesion may provide an important improvement to the assessment of text complexity. In sum, it is important to be able to understand, quantify, and control text complexity for the purposes of: building student skills in reading and understanding increasingly complex text, preparing students for the reading demands of college and the working world, improving summative assessment practices, and reducing variability in formative assessment tools to facilitate better decision making.

Additionally, the Coh-Metrix tool may provide a means to evaluate text beyond those features captured in readability formulas. Coh-Metrix allows for a quantitative evaluation of features of text cohesion, and can support researcher understanding of the

potential contributions of cohesion to text complexity. Consequently, Coh-Metrix can be used as a tool in evaluating the referential cohesion of a passage, and the effects of referential cohesion on reading performance.

CHAPTER III

METHODOLOGY

As outlined in Chapter 1, it was hypothesized that referential cohesion and passage genre have direct and indirect effects on student oral reading fluency rate, accuracy, and passage-specific comprehension. In order to evaluate these hypotheses, this study included two qualitative independent variables, each with two levels – referential cohesion (high/low) and genre (narrative/informational). The study evaluated the effects of these independent variables on three dependent variables – oral reading fluency rate, oral reading fluency accuracy, and passage-specific reading comprehension. Participating students read four passages that were strategically selected to manipulate referential cohesion and genre while tightly controlling readability. Selected passages represented the following conditions: 1) informational text/low cohesion, 2) informational text/high cohesion, 3) narrative text/low cohesion, and 4) narrative text/high cohesion. The study design allowed for evaluation of direct effects of referential cohesion and genre on rate, accuracy, and passage-specific comprehension, as well as interaction effects between referential cohesion and passage genre on dependent variables.

Independent Variables

Two independent variables were manipulated in this study: genre and referential cohesion.

Genre. Passage genre was identified as either narrative or informational. Genre for all passages was determined by the authors of the selected measure and verified by

expert ratings. Genre is defined as a dichotomous qualitative variable, with two levels: narrative and informational. Narrative texts are defined as writing that

conveys experience, either real or imaginary, and uses time as its deep structure. It can be used for many purposes, such as to inform, instruct, persuade, or entertain (Common Core Standards Initiative Appendix A, 2010, p. 23).

Informational texts are texts that convey

information accurately. This kind of writing serves one or more closely related purposes: to increase readers' knowledge of a subject, to help readers better understand a procedure or process, or to provide readers with an enhanced comprehension of a concept. Informational/explanatory writing addresses matters such as types...and components...; size, function, or behavior...; how things work...; and why things happen (Common Core Standards Initiative Appendix A, 2010, p. 23).

Author judgments of genre were evaluated by a panel of graduate student expert reviewers using these definitions. All reviewers have received graduate-level training in school psychology, and have studied early literacy intervention and assessment.

Reviewers were provided with selected passages in a random order and the Common Core Standards Initiative (2010) definitions of narrative and informational text, and asked to label passages as narrative or informational. Reviewers were in 100% agreement with each other and passage authors on genre assignments. See Table 3 for genre definitions and reviewer and author ratings.

Referential cohesion. A researcher-developed referential cohesion composite score was created for this study. The researcher-developed referential cohesion

Table 3.

Expert Reviewer and Passage Author Judgments of Passage Genre.

Passage	Reader 1	Reader 2	Reader 3	Author Judgment
1	Informational	Informational	Informational	Informational
2	Narrative	Narrative	Narrative	Narrative
3	Narrative	Narrative	Narrative	Narrative
4	Informational	Informational	Informational	Informational

Note: Definitions of narrative and informational text were provided from the Common Core Standards Initiative (2010). Narrative texts are defined as texts that convey experience, either real or imaginary, and use time as the structure. They can be used for many purposes, such as to inform, instruct, persuade, or entertain. Informational texts are texts that convey information accurately. These kinds of text serve one or more closely related purposes: to increase readers' knowledge of a subject, to help readers better understand a procedure or a process, or to provide readers with an enhanced comprehension of a concept.

composite score (RCCS) was meant to capture variables that are conceptually related to the construct of referential cohesion as described in the integrated model of cohesion. In order to determine which variables to include in the RCCS, the primary researcher performed a qualitative evaluation of all Coh-Metrix variables to identify variables that can be linked to the referential cohesion devices outlined in the integrated cohesion model. This evaluation was based on existing work using Coh-Metrix to measure referential cohesion (see Hiebert, 2011; Graesser, McNamara, & Kulikowich, 2011) and correspondence to the integrated model of cohesion. A total of five variables were identified. These five variables were determined to be explicitly related to the integrated model of cohesion: adjacent anaphor overlap, adjacent argument overlap, content word overlap, stem overlap, and latent semantic analysis (sentence all). Figure 4 outlines the relations between these variables and referential cohesion, as conceptualized in the

integrated model of cohesion. Each of these variables is described in detail below and summarized in Table 4.

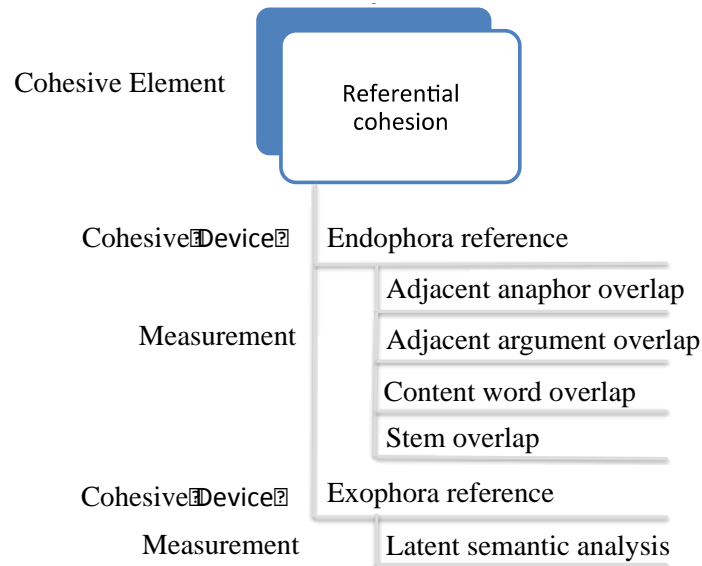


Figure 4. Measurement model of referential cohesion.

Adjacent anaphor overlap. The Coh-Metrix adjacent anaphor overlap variable captures the proportion of anaphor (pronouns that refer to previous nouns) references between adjacent sentences. For example, in the sentences “Jasmine stayed up all night studying for a physics exam. In the morning, she was exhausted,” the anaphor “she” in the second sentence refers to referent “Jasmine.” Adjacent anaphor overlap is a feature of endophora reference because anaphors are co-referent within the text, and do not require the reader to reference information outside of the text.

Adjacent argument overlap. The Coh-Metrix adjacent argument overlap variable captures the proportion of adjacent sentences that share arguments. An argument refers to a noun, pronoun, or noun phrase. Consider the following sentences: “Jimmy’s family went out for ice cream. Jimmy chose chocolate, because it is his favorite flavor.” In both

of these sentences, the noun “Jimmy” is used, linking the sentence through a shared reference. Adjacent argument overlap is a component of endophora reference, because it maintains a continuity of reference within the text itself.

Content word overlap. The Coh-Metrix content word overlap variable captures the proportion of content words that overlap between adjacent sentences. For example, consider these sentences: “The American Civil War was initiated by the secession of several states from the Union. A total of eleven states declared their session and formed the Confederate States of America.” In these sentences, “secession” is a key content word and is represented in both sentences, linking the sentences through content word overlap. Unlike LSA, in which conceptually similar words are activated through shared semantic networks, content word overlap is a feature of endophora reference because it creates explicit linkages within the text. In other words, content word overlap does not require the reader to infer relations between words based on conceptual similarity; instead, these relations are made explicit through continuity of reference within the text.

Stem overlap. The Coh-Metrix stem overlap variable captures the proportion of all sentence pairs in a paragraph that share one or more word stems. In this context, a stem refers to a core morphological element. For example, the words “electricity” and “electrical” share a common morphological element – the word part “electric,” which informs the reader that both words refer to flow of electrical charges. The overlap of the word part “electric” informs the reader that both sentences are referring to the same thing, idea, or concept – they are co-referent. Stem overlap is a feature of endophora reference because co-reference is contained within the text through the repetition of shared morphological elements.

Table 4.

Coh-Matrix Variables Included in the Researcher-Developed Referential Cohesion Composite Score (RCCS).

Variable	Description	Example	Discussion
Adjacent argument overlap	Proportion of adjacent sentences that share common arguments (nouns, pronouns, or noun phrases)	Cell division occurs to reproduce and replace <u>cells</u> . The division of <u>cells</u> with a membrane-bound nucleus and organelles (eucaryotic cells) involves two distinct but overlapping stages, mitosis and cytokinesis.	The word <i>cells</i> overlaps between two adjacent sentences
LSA sentence all	Conceptual similarity of word meanings across all sentences	The field was full of lush, green grass. The horses grazed peacefully. The young children played with kites. The women occasionally looked up, but only occasionally. A warm summer breeze blew and everyone, for once, was almost happy.	The words in the text tend to be thematically related to a pleasant day in an idyllic park scene: <i>green, grass, children, playing, summer, breeze, kites, and happy</i>
Content word overlap	Proportion of content words that overlap between adjacent sentences	One stage of <u>cell</u> division is <u>mitosis</u> . <u>Mitosis</u> occurs to replicate the <u>cell</u> 's genetic material in the nucleus.	The words cell and mitosis are content-specific words that recur across sentences.
Stem overlap	Proportion of all sentence pairs in a paragraph that share one or more word stems (core morphological element)	The <u>division</u> of cells with a membrane-bound nucleus and organelles (eucaryotic cells) involves two distinct but overlapping stages, mitosis and cytokinesis. Mitosis occurs to replicate the cell's genetic material in the nucleus, whereas cytokinesis occurs to <u>divide</u> the gel-like liquid surrounding the cell's nucleus, called cytoplasm.	The word <i>division</i> has a stem overlap with <i>divide</i>
Adjacent anaphor overlap	Proportion of anaphor (pronouns that refer to previous nouns) references between adjacent sentences	There are four distinct <u>phases</u> of mitosis called prophase, metaphase, anaphase, and telophase. These four phases are well known to researchers who can easily observe <u>them</u> with, for example, the simple light microscope.	The pronoun <u>them</u> refers to <u>phases</u> in the previous sentence

Source: McNamara, D.S., Louwse, M.M., Cai, Z., & Graesser, A. (2005).

Latent semantic analysis (sentence all). The Coh-Metrix latent semantic analysis (sentence all) variable captures the conceptual similarity of word meanings across all sentences using a procedure called latent semantic analysis (LSA). Latent semantic analysis is a computer-based method of capturing the semantic similarity of words in the text based on frequent word co-occurrence (Magliano & Millis, 2003). For example, words like “solar system” and “planets” are more likely to co-occur than words like “solar system” and “blueberry;” LSA captures the similarity of words through statistical analysis and provides an overall score between 0 and 1. While latent semantic analysis evaluates text on a semantic level, it can be considered a component of referential cohesion and not lexical accessibility and diversity because it captures the extent to which words in a text relate to one another and activate similar semantic networks, a feature of exophora reference.

Prior work has established a precedent for evaluating referential cohesion using a researcher constructed variable. For example, recent work by Hiebert (2011) used individual Coh-Metrix variables to create a referential cohesion composite. In this work, Hiebert created a referential composite score using argument overlap and stem overlap variables. These variables capture components of the endophora reference domain of referential cohesion, but they do not represent all of the available information that may contribute to referential cohesion – namely, other aspects of endophora reference (such as anaphor reference) and exophora reference. As a result, the RCCS was created for this study to capture a broader range of cohesive devices that contribute to referential cohesion.

Constructing the RCCS. Before constructing the RCCS the researcher evaluated

inter-correlations between the five variables to be included in the composite, as variables that capture a common construct should, in theory, be inter-correlated. Results of these correlations are presented in Table 5.

Table 5.

Inter-Correlations Between Variables Included in the Referential Cohesion Composite Score (Z-Scores).

	Content word overlap	Stem overlap	Adjacent anaphor overlap	Latent semantic analysis
Adjacent argument overlap	.78**	.54**	.24	.50**
Content word overlap		.38**	.20	.55**
Stem overlap			-.43**	.63**
Adjacent anaphor overlap				-.29*

Note: Correlations marked with a * are significant at the $p < .05$ level. Correlations marked with a ** are significant at the $p < .01$ level.

In general, small to modest correlations were found between the variables included in the RCCS. This was to be expected, as each variable captures a small and distinct component of the larger construct of referential cohesion. Some of the individual variables were not correlated; this was also expected, as it is not possible for these variables to occur together. For example, correlations between adjacent argument overlap and adjacent anaphor overlap were expected to be non-significant because anaphors are used in lieu of, rather than in addition to, arguments. In other words, sentence pairs that demonstrate adjacent argument overlap will, by definition, fail to demonstrate adjacent anaphor overlap because arguments are used in place of anaphors.

The first step in creating the RCCS was to set each variable to the same scale of measurement by converting it to a z-score for use in a unit-weighted improper linear

model (Dawes, 1979). The Coh-Metrix program provides raw counts for each variable included in the RCCS, and the metric varies based on what is being calculated (e.g., proportion, frequency, etc.). Using Coh-Metrix analysis output for all considered passages ($N = 29$), means and standard deviations were calculated for each variable included in the RCCS. These means and standard deviations were then used to convert raw scores into z -scores. Once variables were converted to a standard metric, they could be combined to create a composite score. Because there were no hypothesized differences in how each variable contributes to the overall referential cohesion of a passage, all variables were equally weighted by averaging the z -scores together (Dawes, 1979). The resulting score was also converted to a z -score, which became the RCCS. RCCS scores ranged from -1.12 to 3.17. It is unknown how these values relate to other methods of measuring referential cohesion.

While the RCCS was measured quantitatively, referential cohesion was treated as a qualitative variable with two levels: high cohesion (RCCS above the 75th percentile) and low cohesion (RCCS below the 25th percentile). This decision was made to maximize differences in referential cohesion, and to capture how text complexity estimates are used in application. For example, educators may select texts based on the assigned reading level (i.e., 1st grade, 2nd grade, 3rd grade, etc.). While text complexity estimates may vary on a quantitative scale, in practice educators rely on ordinal scales to interpret and apply text complexity information.

Readability

In order to evaluate the unique contribution of referential cohesion on reader comprehension and reading rate and accuracy, readability was held constant across all

passages. All passages considered for inclusion in the study were evaluated using the Lexile® Framework for Reading. The Lexile Framework for assessing text complexity evaluates text on two domains: syntactic and semantic complexity. As in other widely available readability formulas (see Klare, 1974), Lexiles use mean sentence length as a proxy for syntactic complexity. Where Lexiles differ from other readability formulas is the evaluation of semantic complexity; rather than using high-frequency word lists to categorize uniqueness of words, the Lexile measure draws from a corpus of texts containing nearly 600 million words (Lennon & Burdick, 2004). Based on these two variables, texts are assigned a Lexile score ranging from 200 to 1700+, with lower scores indicating higher readability (i.e., lower text complexity) and higher scores indicating lower readability (i.e., higher text complexity).

For this study, passages were selected to meet specific criteria for referential cohesion, readability, and genre. As a result, it was not possible to target passages within a specified readability range (e.g., selecting only highly readable passages). Instead, the pool of passages was first narrowed based on referential cohesion and then genre, and, from the remaining pool, passages with nearly identical readability were selected for study inclusion.

Manipulating Independent Variables: Passage Selection

Passages were strategically selected from the available set of passage probes to allow for the testing of the main effects of referential cohesion and genre and two-way interaction effects between the independent variables. Because the selected population is third grade students, all third grade DIBELS Oral Reading Fluency (DORF) passages were considered for study inclusion. All benchmark and progress monitoring passages

were considered, for a total of 29 passages. From these 29 passages, four passages were selected to test the effects of referential cohesion and genre on reading rate, accuracy, and comprehension using the following procedure:

Measure referential cohesion and identify “low” and “high” cohesion passages. Once the battery of potential passages was identified, all passages were analyzed using the Coh-Metrix tool and the referential cohesion composite score (RCCS) was computed. The 29 passages were then divided into quartiles based on RCCS score in order to identify passages with high and low cohesion. All passages below the 25th percentile ($RCCS < -0.75, n = 7$) were considered for inclusion as “low cohesion passages,” and all passages above the 75th percentile ($RCCS > 0.48, n = 7$) were considered for inclusion as “high cohesion passages.”

Identify passages with similar readability scores. The passages in the DIBELS Next assessment battery were developed with the intent of closely controlling readability. Within each grade level, only texts that represented readability within a specified range were included in the final measure (Powell-Smith et al., 2010). Despite such attempts to control for readability, passages included in the third grade set of oral reading fluency measures range in readability from 640 to 860 on the Lexile scale (according to the Common Core Standards [2012] Lexile to grade correspondences, this range of Lexile scores spans third and fourth grades; using the MetaMetrix [2013] Lexile to grade correspondences, these scores span third through eighth grades). Because this study is designed to evaluate the effects of referential cohesion when readability is held constant, readability scores were examined for each potential passage. From the sample of seven

“low cohesion passages” and seven “high cohesion passages,” the researcher identified passages that were nearly identical in Lexile score.

Identify two passages within each genre. Once high and low cohesion passages with similar readability scores were identified, the researcher identified one high cohesion and one low cohesion text within each genre. A total of four passages were selected, to capture the following conditions: 1) informational text/low cohesion, 2) informational text/high cohesion, 3) narrative text/low cohesion, and 4) narrative text/high cohesion. See Table 6 for detailed information about each of the selected passages.

Table 6.

DIBELS Next Oral Reading Fluency Passages Selected for Study Inclusion.

Condition	Probe	Genre	Lexile	RCCS	Condition
A Woodland Path	Progress Monitoring #7	Narrative	760	0.85	Narrative/High cohesion
Living in Singapore	BOY Benchmark #3	Narrative	750	-1.19	Narrative/Low cohesion
Raising a Calf	MOY Benchmark #2	Informational	790	1.22	Informational/High cohesion
Save the Turtles!	Progress Monitoring #11	Informational	790	-0.81	Informational/Low cohesion

Note: BOY = beginning of year, MOY = middle of year, RCCS = referential cohesion composite score.

Dependent Variables

Three dependent variables were selected for examination: reading rate, accuracy, and passage comprehension. These variables all capture components of oral reading fluency, defined as:

efficient, effective word recognition skills that permit a reader to construct the meaning of text. Fluency is manifested in accurate, rapid, expressive oral reading and is applied during, and makes possible, silent reading comprehension” (Pikulski & Chard, 2005, p. 3).

Dependent variable #1: rate. As noted by Pikulski & Chard (2005), one component of oral reading fluency is reading with sufficient rate. Traditionally rate, as measured by the number of words read correctly in one minute (wcpm), is the primary score obtained in oral reading fluency measures. Reading rate was selected as a dependent variable because: 1) it captures the complex integration of multiple skills necessary for reading with comprehension, and 2) there is evidence that reading rate is a strong indicator of overall reading performance. First, reading rate captures the process of mastering decoding skills to the point of automaticity, a critical component of reading with comprehension as recognized by automaticity, interactive, and reciprocal relationship theories of reading development (Fuchs, Fuchs, Hosp, & Jenkins, 2001). Second, empirical research indicates that measures of oral reading fluency (which determine reading competence largely on wcpm scores) may be more highly correlated with a criterion test of reading comprehension than more direct methods of measuring reading comprehension, such as question answering, retelling, and cloze procedures (Fuchs et al., 2001). Additionally, strong correlations have been found between rate scores and student performance on high-stakes state testing (Wood, 2006; Stage & Jacobsen, 2001).

Dependent variable #2: accuracy. Oral reading accuracy score, expressed as a percentage of words read correctly, captures a second component of Pikulski & Chard’s

(2005) definition of fluency. Accuracy was included as a dependent variable because accurate reading is a critical part of reading proficiency, as accuracy is necessary for comprehension. As noted by Kame'enui and Simmons (2001) "fluency as an index of sheer speed without accuracy is a reckless indicator of processing, cognitive or otherwise. Instead, fluency should always serve to index both accuracy and speed" (p. 206). Consistent with this argument, a measure of accuracy is included in the evaluation of student oral reading fluency.

Dependent variable #3: comprehension. It was hypothesized that passage comprehension is one mechanism by which referential cohesion affects oral reading fluency rate and accuracy. As a result, it was essential that the selected reading comprehension measure capture passage-specific comprehension, rather than global comprehension or verbal reasoning skills. While many tools are available to capture passage-specific reading comprehension, a passage recall task was selected to measure comprehension. The selected task allows for measurement of comprehension at the individual idea unit-level, providing a far greater sample of student responses than a question-based task. It also specifically targets the text and the text's impact on understanding, rather than student-level comprehension construction and integration skills. While these are important and meaningful components of comprehension, understanding the specific comprehension processes students use in reading a text is beyond the scope of this study. Previous work in the area of text cohesion and reading comprehension has used a variety of comprehension measures, such as cloze (e.g., Greenfield, 1999), recall (e.g., Beck, McKeown, Omanson, and Pople, 1984; Beck, McKeown, Sinatra, and Loxterman, 1991; Britton and Gulgoz, 1991; McNamara &

Kintsch, 1996; Vidal-Abarca, Martinez, and Gilabert, 2000), multiple-choice questions (e.g., Britton and Gulgoz, 1991; McNamara & Kintsch, 1996), open-ended questions (e.g., Beck, McKeown, Sinatra, and Loxterman, 1991; McNamara & Kintsch, 1996; Vidal-Abarca, Martinez, and Gilabert, 2000; McNamara, 2011), keyword sorting or association (e.g., Britton and Gulgoz, 1991; McNamara & Kintsch, 1996). In selecting a comprehension measure for this study, it was important that the comprehension measure could be used in combination with a measure of oral reading fluency, as both measures are based on the same passage. This restriction makes it challenging to use cloze or maze procedures, as they are not designed to be used after a student has already read the complete passage to measure oral reading fluency. Multiple choice and open-ended questions can be given after a student completes a one-minute timed read of the passage for fluency, but student performance is largely related to the quality of the questions and, in the case of multiple-choice tasks, response choices. Additionally, comprehension questions generally sample understanding from select portions of the passage; while main idea-type questions may capture whole-passage comprehension, they require student-level comprehension skills that are independent of the task. For example, these types of questions may require a student to use deductive or inductive reasoning skills, which represent general comprehension skills rather than specific understanding of the passage itself. For these reasons, a recall task was selected as the measure of comprehension.

Measures

Oral reading fluency. Student oral reading rate and accuracy were assessed using the passages from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) – Next Edition, a curriculum-based measurement system. One of the measures, DIBELS

Oral Reading Fluency (DORF), reportedly captures a number of components of overall reading proficiency including advanced phonics skills, accurate fluent reading of connected text, and reading comprehension (Good, Kaminski, Dewey, Wallin, Powell-Smith, & Latimer, 2011). According to the DIBELS Next Technical Manual (Good et al., 2011), the standard error of measurement (SEM) for the wcpm score for a single DORF passage in third grade is 11.29. The SEM can be used to compute a confidence interval for an individual test score. In order to calculate a 95% confidence interval, the SEM is multiplied by 1.96. For example, if a student earned a DORF rate score of 92 wcpm, there is 95% confidence that the student's true score lies within the range of 70 to 114 wcpm.

In order to evaluate the effect of referential cohesion on oral reading fluency and comprehension, the DIBELS Oral Reading Fluency (DORF) passages were administered and scored in a non-standardized manner. Students were administered each passage using standardized DIBELS Next directions, but, rather than stop the student after one minute, the examiner allowed the student to read the entire text. The examiner recorded the total time it took for the student to read the passage, and this time was used to calculate an overall words correct per minute (wcpm) rate score (words correct/total time in seconds * 60). Accuracy score was calculated based on performance on the entire passage (words correct/total words in passage * 100). These alternative procedures were used for two reasons. First, text cohesion and readability estimates were calculated based on the entire passage. It is possible that the first minute of the passage (which will vary for each student) is more or less readable or cohesive than the entire passage; allowing students to read the entire passage ensures that readability and referential cohesion ratings

align with the actual text students are exposed to. Second, standardized administration constrains the extent to which the student can access the comprehension measure. Specifically, students that read more text are able to demonstrate comprehension of a greater number of idea units than students who read with a slower rate. Allowing all students to read the entire passage provided all students with exposure to the same amount of content.

In order to assess the validity of these non-standardized procedures, two additional types of oral reading fluency data were collected. First, the primary researcher accessed school-wide easyCBM oral reading fluency data collected as a part of school-wide benchmarking procedures for all students who participated in the study. These data were collected during the same two-week period as study data collection. While passage development and equating procedures are different than those used by the authors of the passages used in the study, easyCBM oral reading fluency probes follow standardized CBM procedures to provide a measure of oral reading fluency rate (words read correctly per minute; wcpm). Second, during data collection examiners recorded student scores for the first minute of administration in addition to scores for the entire passage (referred to as “first minute” scores). Descriptive statistics for all passages can be found in Table 7.

An examination of correlations between the three types of scores indicates that the non-standardized DIBELS Next wcpm scores (referred to as “pro-rated” scores) correlate strongly with easyCBM and first minute DIBELS Next wcpm scores for all passages ($r = .91$ to $.98$). These correlations suggest that the non-standardized, pro-rated procedure did not compromise the validity of the rate measure, and that the results and

Table 7.

Descriptive Statistics for easyCBM Benchmark and Study Passage Rate Scores (First Minute and Pro-Rated Whole Passage).

Passage	<i>N</i>	<i>M</i>	<i>SD</i>
easyCBM Benchmark	74	118.12	40.50
Narrative/High Referential Cohesion Pro-Rated	74	94.55	38.05
Narrative/High Referential Cohesion First Minute	74	96.54	39.68
Narrative/Low Referential Cohesion Pro-Rated	74	88.28	33.16
Narrative/Low Referential Cohesion First Minute	74	84.15	34.36
Informational/High Referential Cohesion Pro-Rated	74	90.97	33.71
Informational/High Referential Cohesion First Minute	74	85.72	32.06
Informational/Low Referential Cohesion Pro-Rated	74	89.09	32.81
Informational/Low Referential Cohesion First Minute	74	91.12	33.16

conclusions are likely to be applicable to common educational practice for progress monitoring. See Table 8 for correlations for all passages.

Passage recall. Student passage-level comprehension was measured using a passage recall task. This task is based on the work of McMaster and colleagues (McMaster et al., 2012), who adapted the coding scheme of van den Broek and colleagues (e.g., Kendeou & van den Broek, 2005; Linderholm & van den Broek, 2002) in order to capture student recall at the idea-unit level. In this type of recall, students are asked to retell the passage and responses are coded based on how closely students captured the meaning or gist of each idea unit in the text.

In preparation for data collection, each original passage was parsed into individual idea units. An idea unit is defined as a distinct, identifiable, and meaningful idea, which

Table 8.

Correlations Between easyCBM Benchmark and Study Passage Rate Scores (First Minute and Pro-Rated Whole Passage).

	N-HRC Pro- Rated	N-HRC First Min	N-LRC Pro- Rated	N-LRC First Min	I-HRC Pro- Rated	I-HRC First Min	I-LRC Pro- Rated	I-LRC First Min
easyCBM	.94	.91	.95	.93	.94	.94	.94	.91
N-HRC Pro-Rated		.97	.97	.96	.98	.96	.97	.94
N-HRC First Min			.95	.95	.96	.95	.96	.94
N-LRC Pro-Rated				.98	.97	.96	.97	.95
N-LRC First Min					.96	.95	.96	.94
I-HRC Pro-Rated						.98	.97	.95
I-HRC First Min							.96	.95
I-LRC Pro-Rated								.97

Note: N-HRC = Narrative/high referential cohesion passage, N-LC = Narrative/low referential cohesion passage, I-HRC = Information/high referential cohesion passage, I-LRC = Informational/low referential cohesion passage. All correlations are significant at the $p < .01$ level.

generally includes a subject and a verb and constitutes an independent or dependent clause. For example, consider the following sentence from a selected passage “Of the seven species of sea turtles, the largest is the leatherback.” This sentence was parsed into two idea units because there are two distinct thoughts expressed in the sentence – 1) “of the seven species of sea turtles” (idea: there are seven species of sea turtles), and 2) “the largest is the leatherback” (idea: the largest species of sea turtle is the leatherback). This definition allows for researcher judgment in important main ideas and allows the researcher to capture the specific information of interest, and is consistent with previous work by McMaster (2012). Each passage was parsed into 34-37 individual idea units.

The recall task was administered immediately after each oral reading fluency passage. Students were presented with study-standardized recall directions. All student recalls were recorded for transcription and coding. Student recalls were untimed. General prompts were provided until students indicated that they could not remember any additional information about the text (e.g., Lynch & van den Broek, 2007)

After data collection concluded, student recalls were parsed into individual idea units and compared to the original text idea units. Based on this comparison, recalled idea units were coded as: 1) conservative, 2) liberal, 3) no match-consistent, or 4) no match-inconsistent. These codes were developed by McMaster and colleagues (McMaster et al., 2012), and definitions are consistent with those provided by the original authors. For the purpose of this research, one code was omitted (highly connected). This code was designed to capture the number of causal connectives in the student's recall. Because causal consistency (a feature of grammatical cohesion) was not a variable of interest for this research, it was omitted from the design. Descriptions of each code follow, and examples of these codes as applied to actual student responses can be found in Table 9.

Conservative. Conservative recalled idea units are literal or near-literal retellings of the targeted idea unit. A conservative response accurately captures the meaning or gist of the idea unit and includes most or all of the words in the original text. A conservative response also captures all important components of the original idea unit (e.g., all characters or actions).

Liberal. Liberal recalled idea units are non-literal retellings of the targeted idea unit. A liberal response somewhat captures the primary meaning of the idea unit, but

may be summarized in the reader's own words. Additionally, a liberal response may omit a detail or important component from the original idea unit.

No match-consistent. No match-consistent recalled idea units are retellings that cannot be matched directly to an idea unit in the text, but represent a logical or valid inference based on the text. No match-consistent responses are consistent with the text meaning but go beyond what is included in the original text. The inclusion of this code in the coding scheme allows the comprehension measure to capture readers who have gone beyond the text to form a mental representation of the passage meaning.

No match-inconsistent. No match-inconsistent recalled idea units do not match directly with an idea unit and are inconsistent with the meaning of the text. These responses may be incorrect recall of text information, student opinion, off-track responses, etc.

Each retell was assigned a score for each of the four codes: 1) total number of conservative responses, 2) total number of liberal responses, 3) total number of no match-consistent responses, and 4) total number of no match-inconsistent responses. The total number of conservative responses, liberal responses, and no-match consistent responses were added together to obtain a total number of consistent recall responses. This score was then divided into the total number of idea units from the original text, in order to obtain a proportion of consistent (or "correct") responses. It was possible for students to earn a comprehension score that exceeded one (i.e., student provided conservative or liberal responses for all idea units in the text and provided no-match consistent responses); however, this did not occur. This comprehension score is slightly different from that used by McMaster and colleagues (McMaster et al., 2012). In McMaster's

Table 9.

Sample Coding of Student Responses to the Passage Retell Task.

Original Idea Unit	Student Response	Code	Rationale
The largest is the leatherback.	Leatherbacks are the biggest sea turtles.	Conservative	The student’s response is a near literal retelling of the original text. The response captured the central idea of the idea unit – that leatherbacks are the largest – and included the implicit co-referent – leatherbacks are the largest <i>of the sea turtles</i> .
Other types of sea turtles are not able to do this.	And other turtles can’t really do that.	Conservative	While this student used his/her own words, this response captures all of the important parts of the original idea unit – that there are other types of turtles, and that they are not able to do something.
One thing Nell and her family had to get used to was the rain.	They had to get used to the rain	Conservative	The student’s response includes all of the important features of the original text – that the subject is Nell and her family (captured by the use of “they”) and that they had to get used to the rain. It also includes most of the words from the original idea unit, making it a near-literal retelling.
In the clearing was the most beautiful waterfall they had ever seen.	They found a waterfall	Liberal	The student’s response captured the main idea of the original idea unit, which is that the children found a waterfall. However, the response is missing key details, such as the waterfall being the most beautiful that the children had ever seen.
The whole family moved	She moved somewhere	Liberal	The student’s response captured the gist of the idea unit (someone moved) but failed to include a key detail – that it was the <i>whole family</i> and not just the protagonist that moved. Inclusion of this detail would make this response Conservative.
They are called leatherbacks because they have a softer, more flexible shell than other turtles.	They have much softer and more flexible shells than other turtles.	Liberal	This response captured the central idea of the original idea unit – that the leatherback’s shell is softer and more flexible than that of other turtles. However, it omits a key detail – that this shell is why leatherbacks were given that name. This omission makes this a liberal response (it captures the gist, but excludes some key words or details from the original).

Table 9 continued

Original Idea Unit	Sample Response	Code	Rationale
	You must not throw plastic bags or anything in the ocean	No Match-Consistent	This response cannot be matched to an idea unit in the original text, as the text never explicitly stated that people should not throw trash into the ocean. However, the text did state that plastic bags are harmful to sea turtles, and that people are beginning to recycle and throw away fewer plastic bags. It would be a logical inference to make that people should not throw plastic bags into the ocean (because they are harmful to turtles).
	Nell wanted an ice cream cone,	No Match-Consistent	The text states that the protagonist (Nell) stopped and stared at the snow cones, and then her dad bought her one. While her mental state is not stated explicitly, it would be reasonable to deduce that she was staring at the snow cone and got one because she wanted one.
	They like hiking.	No Match-Consistent	The original text does not contain an idea unit in which it is explicitly stated that the characters enjoy hiking. However, the text states that the characters hike every day, spend the whole day exploring, and are excited when they find a new path. It is reasonable to infer that the characters like to hike.
	You never put dresses on turtles.	No Match-Inconsistent	This response cannot be matched to an idea unit in the original text, making it a No Match response. It is inconsistent with the original text because it is not a logical and reasonable inference that could be made from the text.
	Who would name a girl Nell?	No Match-Inconsistent	This response cannot be matched to an idea unit in the text. It is an inconsistent response because it is a non-sequitur.
	Well they lived in a small cottage	No Match-Inconsistent	The original text does not provide any information about where the characters live, nor it is suggested that their home is small. It is not an unreasonable inference, but is also not supported in any way by the text, making it an inconsistent response.

work, the number of conservative, liberal, highly connected (code not used in this study), no match-consistent, and no match-inconsistent codes were added together and divided by the total number of idea units in the story. This procedure was not used in the present study because: 1) the highly connected code was not used in this study, and 2) the author determined that no match-inconsistent responses do not indicate passage-specific comprehension, and consequently should not be counted toward the student's comprehension score in this study.

Participants

Participants were recruited from two public elementary schools in the Pacific Northwest. Neither school currently uses DIBELS Next Oral Reading Fluency (DORF) passages for screening or progress monitoring. All third grade students at each participating school were invited to participate through an open recruitment letter. A total of 117 students were invited to participate. The parents/guardians of 14 students did not provide consent to participate. Consent forms were not returned for 12 students. Thus, consent to participate was provided for 91 students. Of these students, one was no longer enrolled at the school when testing began and was consequently not assessed. Additionally, seven of the 91 students with consent but did not participate in the study due to absences or scheduling conflicts during the testing window. Consequently, 83 students participated in the study. Of these 83 students, 74 had complete data (rate, accuracy, and comprehension scores for all four passages). Incomplete data was due to passage spoilage ($n = 2$) or student request to discontinue testing ($n = 7$). All analyses included only students with complete data ($n = 74$). Third grade students were selected

because, by third grade, students should have developed enough reading skills to be able to complete the tasks and show meaningful variability in reading competence.

Procedure

Data collector training. Data were collected by graduate students in the special education and clinical services department at the University of Oregon. All data collectors reported having some prior training in DIBELS Next administration. In addition, all data collectors were required to attend a training session in administration and scoring of oral reading fluency passages. This training was led by the primary researcher, who has attended DIBELS Next Essentials and Mentor trainings and is a member of the DIBELS Mentor Network. The training session included background on the measure and its use, review of administration and scoring procedures, and opportunities to practice scoring oral reading fluency probes. At the conclusion of the training, examiner inter-rater agreement data were collected based on live administration of passages to the trainer. Data collector scores were compared to a master key, which was developed by the trainer. All data collectors achieved at least 90% inter-rater agreement with the master key (range: 97% to 98% inter-rater agreement). One data collector chose to complete a second inter-rater agreement check for extra practice. Inter-rater agreement was calculated by dividing the number of items (words) in agreement (correct or incorrect) by the total number of items (words) in the passage, for a percent agreement score.

Data collection. All participating students were asked to read all four selected DORF passages. Passages were presented in random order ($n_{\text{conditions}} = 24$), and examiners and students were blind to passage condition. All passages were administered

using standardized study directions. A discontinue rule was included in the standardized procedures for students who could not read any correct words in the first line; however, no students met criteria for implementation of the discontinue rule. After each passage, students completed the passage recall task using researcher-developed directions. Passage recalls were audio recorded to allow for transcription and idea-unit level coding. Administration time for all four passages was approximately 30 minutes per student. Efforts were made to restrict testing to a single testing session; however, due to variables outside of the researcher's control (e.g., unanticipated interruption, testing taking longer than expected), some students were tested across two sessions. All students were assessed during the same two-week time period in January of 2013.

Inter-rater agreement data were collected on 20% of the final sample (complete data only, $n = 15$). For each examiner, inter-rater agreement was collected for 17% to 21% of students tested. Inter-rater agreement data were collected by comparing item-level scores (words scored as correct or incorrect) for the entire passage as scored by the examiner and a shadow scorer (primary researcher). Inter-rater agreement ranged from 95% to 100% agreement.

Coding of passage recalls. After all data were collected, the audio recordings of the passage retells were transcribed by a professional transcription company. Upon receipt of written transcription of passage recalls, coding of responses began. All passage recalls were coded by the primary investigator by parsing the student recall into idea units and assigning a code to the idea unit based on correspondence with the original text. Based on these codes, each retell was assigned frequency scores for each type of code (conservative, liberal, no match-consistent, no match-inconsistent). The total number of

consistent responses (conservative, liberal, no match-consistent) was divided by the total number of idea units in the original passage to obtain the comprehension score.

Participant Incentives

As a thank you for participating in the study, teachers of participating classrooms ($n = 5$) were given gift cards to a local bookstore, to be used to purchase curricular and other materials for the classroom. These materials were intended to benefit the entire classroom, not just the students that participated in the study. Funding for these gift cards was provided by the research department of the participating school district.

Summary

This study evaluated the effects of referential cohesion and passage genre on oral reading fluency rate, oral reading fluency accuracy, and passage-specific comprehension. This research utilized an experimental, within-subjects, repeated measures design with two qualitative independent variables and three quantitative dependent variables. Strengths of the design include the control of passage readability, the development of a referential cohesion composite which includes all components of referential cohesion as outlined in the integrated model of cohesion, and the repeated measures design.

One notable strength of this design is the passage-specific reading comprehension measure. This measure was carefully selected to capture the reader's understanding of specific elements of each individual passage, rather than a student's global reasoning or inferencing skills. Unlike other measures of reading comprehension, which may measure how much a student recalls or student comprehension of specific features of the text, this recall task captures the breadth of student comprehension of the text (by measuring

comprehension of the entire passage) and the depth of student understanding (by capturing literal and non-literal responses as well as logical inferences).

CHAPTER IV

RESULTS

The final design included two qualitative, within-subjects independent variables with two levels, which were analyzed using two-way analysis of variance (ANOVA) with dependent observations. Three univariate ANOVAs were performed, one for each quantitative dependent variable. Analysis allowed for evaluation of main independent variable effects (genre and referential cohesion) and interaction effects. Analyses evaluated the following research questions:

1. When readability is held constant, do students read more words correctly per minute on passages with higher referential cohesion than passages with lower referential cohesion?
2. When readability is held constant, do students read passages with higher referential cohesion with greater accuracy than passages with lower referential cohesion?
3. When readability is held constant, do students perform better on a measure of passage-specific reading comprehension for passages with higher referential cohesion than passages with lower referential cohesion?
4. When readability and referential cohesion are held constant, do students read more correct words per minute on narrative texts than informational?
5. When readability and referential cohesion are held constant, do students read narrative texts with greater accuracy than informational texts?
6. When readability and referential cohesion are held constant, do students perform better on a measure of passage-specific reading comprehension on narrative texts

than informational?

7. If differences in oral reading performance are noted on high and low cohesion passages (questions 1, 2, and 3), do the effects depend on whether the text is narrative or informational?

Characteristics of the Invited Sample

A total of 117 students across five third grade classrooms in two school schools were invited to participate in the study, though only 116 were still enrolled at the time of winter benchmark assessment. Forty-five students attended School 1, and 71 students attended School 2. All invited students were administered the easyCBM winter benchmark by school personnel as a part of the schools' universal screening processes. Scores are used by school personnel to identify a student's level of risk for future reading failure (cut points determined by the easyCBM authors and participating school district). Students falling in the "low risk" range performed at or above the 50th percentile on national norms, and are considered to have a low risk of future reading failure. Students falling in the "some risk" range performed between the 20th-49th percentiles on national norms, and are considered to have some risk of future reading failure without strategic reading intervention. Students falling in the "high risk" range performed below the 20th percentile on national norms, and are considered to be at a high risk for future reading failure without intensive intervention. At School 1, 22% of all third grade students scored in the "high risk" range on the easyCBM measure of oral reading fluency, 28% of students scored in the "some risk" range, and 50% of students scored in the "low risk range." At School 2, 24% of all third grade students scored in the "high risk" range, 41% of students scored in the "some risk" range, and 35% of students scored in the "low risk"

range. Within the easyCBM system, a typical school should have 50% of students in the low risk range, as the goal is set based on the 50th percentile. Based on this context, the performance of students at School 1 is consistent with other schools using the easyCBM system. School 2, on the other hand, is not consistent with other schools using easyCBM, as only 35% of students fell in the low risk range. Consequently, School 2 may represent a lower performing school system than other schools using easyCBM.

In addition, statewide assessment data provide information about school functioning and context. For participating schools, the most recent statewide assessment data available to the public are from the 2010-2011 school year. At School 1, 85% of third grade students met or exceeded the standard on the state assessment in reading. At School 2, 91% of third grade students met or exceeded the standard. At the district level, 81% of students in grades 3-5 met or exceeded the standard. Consequently, participating schools represent slightly higher achievement on the state assessment than the district average.

Characteristics of the Actual Sample

All participating students were tested on the easyCBM Passage Reading Fluency (PRF) measure during the same two weeks of study data collection as a part of the school's benchmarking process. These easyCBM PRF scores were used to: 1) validate the use of non-standardized scoring procedures for DIBELS Next Oral Reading Fluency, and 2) better understand the skill level of participating students. First, Pearson correlation coefficients were computed for passage rate scores (pro-rated whole passage rate and first minute only rate) and easyCBM benchmark rate scores. The strength and significance of these correlation coefficients supports the validity of the pro-rated rate

score, which was used for all subsequent analyses. Second, easyCBM PRF scores were sorted based on level of risk in order to describe the skill level of participating students. Of the 74 students included in analysis, 13 students performed in the high risk range on the winter easyCBM PRF assessment, which represents 18% of the sample. Thirty-two students performed in the some risk range on easyCBM PRF, which represents 43% of the sample. Finally, 29 students fell in the low risk range, which represents 39% of the sample.

Additionally, these easyCBM PRF scores can be used to compare the final sample used for analysis with the sample of students that were excluded from analysis due to incomplete data. A total of nine students were excluded from the final sample due to incomplete data. This is approximately 11% of the students that were tested. Of these nine students, five have easyCBM PRF scores in the high risk range (56%), two had easyCBM PRF scores in the some risk range (22%), and two had easyCBM PRF scores in the low risk range (22%). Compared to the sample of students with complete data, a greater percentage of students with incomplete data fell in the high risk range (18% of complete sample, 56% of incomplete sample). Accordingly, the complete sample included a greater percentage of students performing in the some risk (43% of complete sample, 22% of incomplete sample) and low risk (29% of complete sample, 22% of incomplete sample) ranges than excluded students. This suggests that data were not missing at random, and that the final sample used for analysis may underrepresent low-performing students.

Data Transformations

Two of the dependent variables, oral reading fluency accuracy and passage-specific comprehension, were measured using counts of correct or appropriate responses divided by the total possible number of responses, resulting in a proportion. Because these proportions were derived from counts, the homogeneity of variance assumption is violated. Additionally, oral reading fluency accuracy scores were negatively skewed, violating the assumption of normality (range across all passages: 0.80 to 1.00). In order to make the data better fit the assumptions of ANOVA, these scores were transformed using the arcsine square root transformation (McDonald, 2009). The arcsine square root transformation is appropriate for these scores as both accuracy scores and passage-specific comprehension scores were expressed as proportions and were constrained between the range of 0 and 1. These transformed values were then used for all ANOVAs. Descriptive statistics reported in Table 10 and Table 11 and scores presented in graphs were back transformed to proportion scores.

Descriptive Statistics

Before exploring evidence related to research questions, descriptive statistics were computed for each variable of interest (rate, accuracy, comprehension) for each passage. See Table 10 for descriptive data for the entire sample. Additionally, descriptives are provided by student skill level (determined by easyCBM risk level on the Passage Reading Fluency measure) in Table 11.

As expected, students falling the low risk range on easyCBM earned higher rate and accuracy scores on study passages than students falling the some risk range, who earned higher scores than students falling in the high risk range. Differences in rate and

accuracy scores across groups were consistent across all passages. While differences are noted in comprehension scores across skill levels, standard deviations indicate that these may not be meaningful differences. Further analysis by skill level was not completed due to the small sample size of each group.

Table 10.

Descriptive Statistics for Rate (Pro-Rated Whole Passage), Accuracy, and Comprehension for all Passages Included in Study.

Passage	Rate		Accuracy		Comprehension	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Narrative/High Referential Cohesion	94.55	38.05	0.96	0.01	0.23	0.02
Narrative/Low Referential Cohesion	88.28	33.16	0.95	0.01	0.24	0.03
Informational/High Referential Cohesion	90.97	33.71	0.95	0.01	0.17	0.02
Informational/Low Referential Cohesion	89.09	32.81	0.96	0.01	0.18	0.02

Note: *N* for all passages was 74. Accuracy and comprehension mean scores are expressed as proportions. Rate score represents words correct per minute based on pro-rated, whole passage reading.

Intercorrelations

In order to better understand relations between variables, intercorrelations for all dependent variables are reported in Table 12.

These correlations indicate that, with the exception of the narrative/low referential cohesion accuracy and narrative/low referential cohesion comprehension scores, all scores are significantly correlated. This suggests that the measures may all be capturing a

related construct or constructs. Correlations between rate scores ranged from .97-.98, indicating strong alternate form reliability. Correlations between comprehension scores were lower but still fairly strong, ranging from .57-.66. This indicates good alternate form reliability, though not as strong as rate. Correlations between accuracy scores ranged from .69-.75, indicating strong alternate form reliability across passages. Across Table 11.

Descriptive Statistics by Risk Level for Rate (Pro-Rated Whole Passage), Accuracy, and Comprehension for all Passages Included in Study.

Passage	Low Risk			Some Risk			High Risk		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Narrative/High Referential Cohesion									
Rate	29	131.17	29.19	32	75.56	15.68	13	49.77	13.08
Accuracy	29	0.98	0.00	32	0.96	0.00	13	0.92	0.01
Comprehension	29	0.28	0.02	32	0.20	0.02	13	0.20	0.01
Narrative/Low Referential Cohesion									
Rate	29	120.55	24.42	32	74.88	14.14	13	49.31	11.85
Accuracy	29	0.98	0.00	32	0.95	0.00	13	0.90	0.01
Comprehension	29	0.30	0.04	32	0.21	0.04	13	0.19	0.02
Informational/High Referential Cohesion									
Rate	29	123.34	24.73	32	78.75	13.61	13	48.85	12.40
Accuracy	29	0.97	0.01	32	0.95	0.00	13	0.91	0.01
Comprehension	29	0.22	0.01	32	0.14	0.02	13	0.14	0.01
Informational/Low Referential Cohesion									
Rate	29	120.31	24.19	32	77.25	13.73	13	48.62	13.85
Accuracy	29	0.98	0.00	32	0.95	0.00	13	0.91	0.01
Comprehension	29	0.22	0.03	32	0.17	0.02	13	0.14	0.01

Note: Accuracy and comprehension mean scores are expressed as proportions. Rate score represents words correct per minute based on pro-rated, whole passage reading. Risk levels determined by performance on winter easyCBM Passage Reading Fluency benchmark.

score types, rate and accuracy scores were moderately correlated, ranging from .58-.69. This is to be expected, as poor accuracy would affect a reader's fluency score; however, strong accuracy alone does not insure a high rate score. Correlations between comprehension and rate (.25-.35) and comprehension and accuracy (.24-.33) were more

Table 12.

Intercorrelations Between Oral Reading Fluency Rate, Oral Reading Fluency Accuracy, and Passage-Specific Comprehension Scores for All Measures.

	N-LRC Rate	I-HRC Rate	I-LRC Rate	N-HRC Comp	N-LRC Comp	I-HRC Comp	I-LRC Comp	N-HRC Acc	N-LRC Acc	I-HRC Acc	I-LRC Acc
N-HRC Rate	.97**	.98**	.97**	.27*	.32**	.34**	.33**	.62**	.69**	.62**	.66**
N-LRC Rate		.97**	.97**	.27*	.31**	.35**	.34**	.58**	.69**	.58**	.66**
I-HRC Rate			.97**	.28*	.30*	.32**	.34**	.58**	.67**	.64**	.64**
I-LRC Rate				.25*	.28*	.32**	.31**	.58**	.68**	.58**	.69**
N-HRC Comp					.59**	.66**	.57**	.27*	.24*	.31**	.26*
N-LRC Comp						.59**	.59**	.28*	.20	.29*	.25*
I-HRC Comp							.52**	.29*	.26*	.33**	.32**
I-LRC Comp								.31**	.32**	.25*	.33**
N-HRC Acc									.75**	.69**	.70**
N-LRC Acc										.58**	.80**
N-HRC Acc											.59**

Note: N-HRC = Narrative/high referential cohesion passage, N-LC = Narrative/low referential cohesion passage, I-HRC = Information/high referential cohesion passage, I-LRC = Informational/low referential cohesion passage, Comp = Comprehension Score, Acc = Accuracy Score. Correlations flagged with * are significant at the $p < .05$ level. Correlations flagged with ** are significant at the $p < .01$ level.

modest.

Oral Reading Fluency Rate

In order to evaluate research questions 1, 4, and 7, a two-way ANOVA with dependent observations was performed with oral reading fluency rate (pro-rated, whole passage) as the dependent variable. It was hypothesized that students would read more correct words per minute on passages with high referential cohesion than passages with low referential cohesion. It was also hypothesized that referential cohesion and genre would interact; however, the nature of this interaction was not hypothesized. There was a significant interaction between genre and referential cohesion on oral reading fluency rate, $F(1, 73) = 10.80, p < .05$. See Table 13 for the ANOVA summary table.

Table 13.

Two-Way, Within-Subjects Analysis of Variance Summary Table for the Effect of Genre and Cohesion on Oral Reading Fluency Rate

<i>Source</i>	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Genre	1	141.98	141.98	3.54
Genre*Subject	73	2927.77	40.11	
Cohesion	1	1228.41	1228.41	27.33*
Cohesion*Subject	73	3281.34	44.95	
Genre*Cohesion	1	356.84	356.84	10.80*
Genre * Cohesion * Subject	73	2211.91	33.04	

Note: *F* values marked with a * are significant at the $p < 0.05$ level.

This interaction effect was further evaluated with pairwise comparisons using the Bonferroni procedure to control family-wise Type I error at .05. Results indicate that rate

scores were significantly higher for the high cohesion narrative text ($M = 94.55$, $SD = 38.05$) than the low cohesion narrative text ($M = 88.28$, $SD = 33.16$). The effect size for this comparison is considered very small, $d = 0.18$, based on Cohen's convention (1988, p. 49, equation 2.3.8). Cohen's d was selected to measure effect size because it provides information about the magnitude of the effect, which can be used to interpret the practical significance of the findings. However, it is important to note that Cohen's d may underestimate the strength of the effect for a power analysis as it does not take into consideration the correlation between the measures. The results of the second pairwise comparison indicated that rate scores were significantly higher for the high cohesion narrative text ($M = 94.55$, $SD = 38.05$) than the high cohesion informational text ($M = 90.97$, $SD = 33.71$). The effect size for this comparison is also considered very small, $d = 0.10$ (Cohen, 1988). The pairwise comparison of high cohesion informational text and low cohesion informational text was non-significant, as was the pairwise comparison of low cohesion narrative text and low cohesion informational text. See Figure 5 for illustration of the referential cohesion differences by genre.

This interaction indicates that, when referential cohesion is high, rate is higher on narrative passages ($M = 94.55$, $SD = 38.05$) than informational passages ($M = 90.97$, $SD = 33.71$). Follow up pairwise comparisons indicate that there is not a significant effect of referential cohesion on informational text reading rate. These findings suggest that referential cohesion may provide greater support for student oral reading fluency rate for narrative passages, but less support for informational passages. Conversely, for passages with low cohesion, informational passages are read at about the same rate as narrative

passages. For passages with high cohesion, informational passages are read at a lower rate.

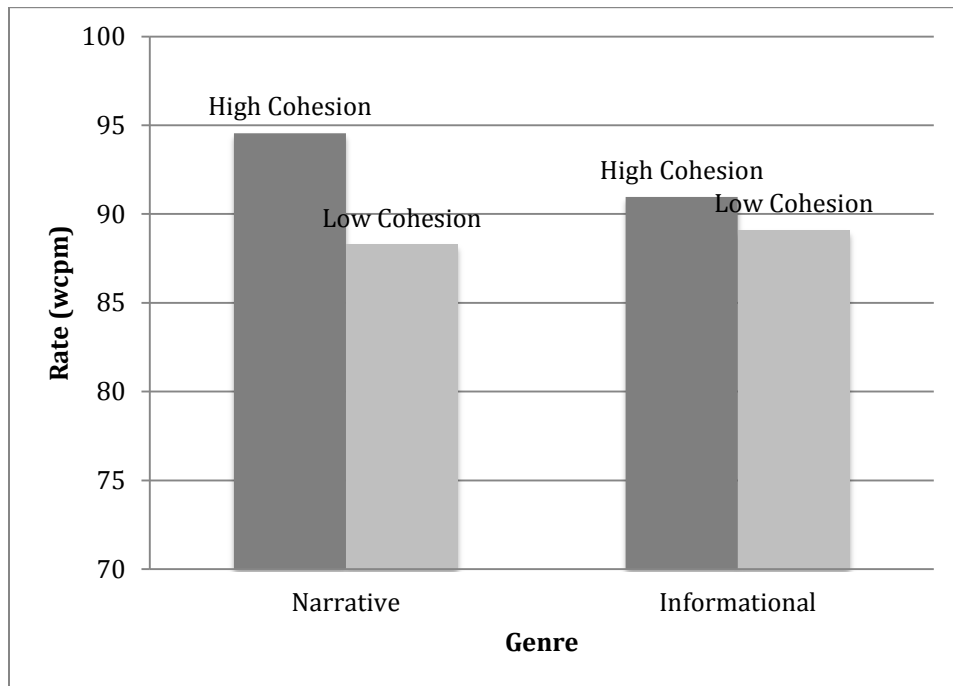


Figure 5. Pairwise comparisons of interaction effects between referential cohesion and genre on oral reading fluency rate.

Oral Reading Fluency Accuracy

In order to evaluate research questions 2, 5, and 7, a two-way ANOVA with dependent observations was performed with oral reading fluency accuracy (based on the entire passage) as the dependent variable. It was hypothesized that students would read passages with high referential cohesion with greater accuracy than passages with low referential cohesion. It was also hypothesized that referential cohesion and genre would interact; however, the nature of this interaction was not hypothesized. The two-way ANOVA yielded a significant interaction between genre and cohesion on oral reading

fluency accuracy, $F(1, 73) = 16.19, p < .05$. See Table 14 for the ANOVA summary table.

Table 14.

Two-Way, Within-Subjects Analysis of Variance Summary Table for the Effect of Genre and Cohesion on Oral Reading Fluency Accuracy.

<i>Source</i>	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Genre	1	0.01	0.01	2.83
Genre*Subject	73	0.15	0.00	
Cohesion	1	0.00	0.00	0.05
Cohesion*Subject	73	0.23	0.00	
Genre*Cohesion	1	0.03	0.03	16.19*
Genre * Cohesion * Subject	73	0.12	0.00	

Note: *F* values marked with a * are significant at the $p < 0.05$ level.

This interaction effect was further evaluated with pairwise comparisons using the Bonferroni procedure to control family-wise Type I error at .05. Results indicate that accuracy scores were significantly higher for the high cohesion narrative text ($M = .94, SD = .01$) than the low cohesion narrative text ($M = .95, SD = .01$). The effect size for this comparison is considered small, $d = 0.25$ (Cohen, 1988). Additionally, accuracy scores were significantly higher for the high cohesion narrative text ($M = .96, SD = .01$) than the high cohesion informational text ($M = .95, SD = .01$). The effect size for this comparison is also considered small, $d = 0.33$ (Cohen, 1988). The pairwise comparison of high cohesion informational text and low cohesion informational text was non-significant, as was the pairwise comparison of low cohesion narrative text and low cohesion informational text. See Figure 6 for illustration of the referential cohesion

differences by genre.

As with rate, these findings indicate that high referential cohesion was related to

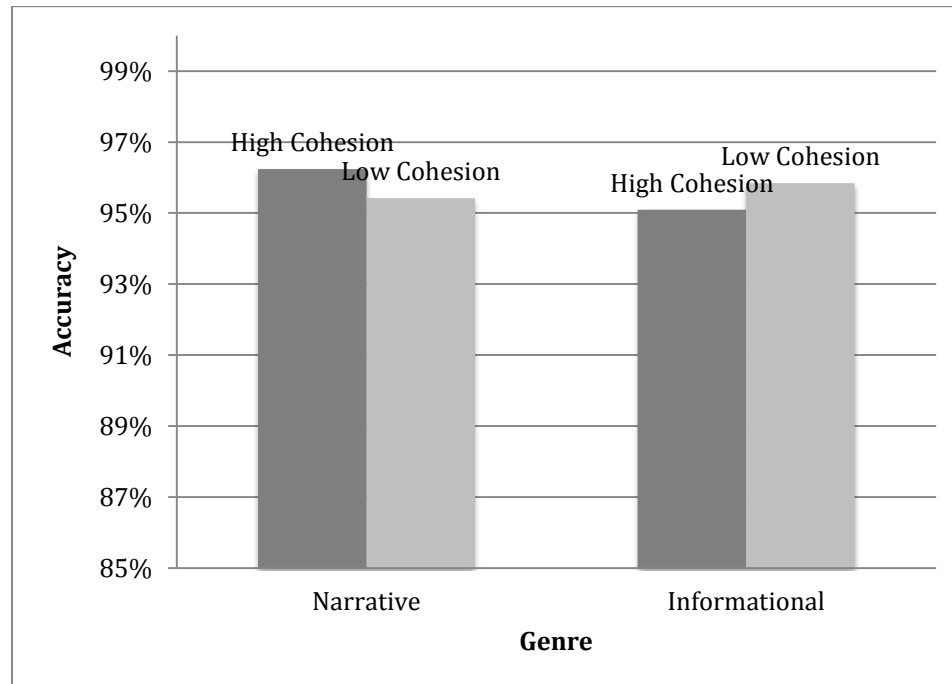


Figure 6. Pairwise comparisons of interaction effects between referential cohesion and genre on oral reading fluency accuracy.

greater accuracy for narrative texts ($M = .96$, $SD = .01$), while low referential cohesion was related to greater accuracy for informational texts ($M = .96$, $SD = .01$). This means that students read narrative passages with high referential cohesion with a greater degree of accuracy than narrative passages with low referential cohesion. Pairwise comparisons indicate that there was no significant effect of referential cohesion on informational text accuracy. These findings suggest that referential cohesion supports greater accuracy of oral reading fluency for narrative passages but not informational passages.

Passage-Specific Reading Comprehension

In order to evaluate research questions 3, 6, and 7, a two-way ANOVA with dependent observations was performed with the comprehension score (proportion of consistent responses) as the dependent variable. It was hypothesized that students would earn higher comprehension scores on passages with high referential cohesion than passages with low referential cohesion. It was also hypothesized that referential cohesion and genre would interact; however, the nature of this interaction was not hypothesized. No hypotheses about main effects for genre were made. Surprisingly, the interaction effect between referential cohesion and genre on passage-specific comprehension was non-significant. An evaluation of main effects indicated that the main effect of referential cohesion was non-significant. See Table 15 for the ANOVA summary table.

Table 15.

Two-Way, Within-Subjects Analysis of Variance Summary Table for the Effect of Genre and Cohesion on Passage-Specific Reading Comprehension

<i>Source</i>	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Genre	1	0.40	0.40	44.61*
Genre*Subject	73	0.65	0.01	
Cohesion	1	0.03	0.03	2.08
Cohesion*Subject	73	0.87	0.01	
Genre*Cohesion	1	0.00	0.00	0.02
Genre * Cohesion * Subject	73	0.71	0.01	

Note: *F* values marked with a * are significant at the $p < 0.05$ level.

This finding indicates that referential cohesion as measured in this study is not related to passage-specific comprehension. There was a significant main effect for genre

on passage-specific reading comprehension, $F(1, 73) = 44.61, p < .05$. Passage-specific comprehension scores were higher on narrative passages than informational passages. These findings indicate that students demonstrated significantly better comprehension of narrative passages ($M = 0.24, SD = 0.02$) than informational passages ($M = 0.18, SD = 0.01$). The effect size for this comparison is considered medium, $d = 0.55$, based on Cohen's convention (1988). See Figure 7 for illustration of the main effect of genre on passage-specific comprehension (reported means are transformed back from the arcsine square root transformation used for analysis).

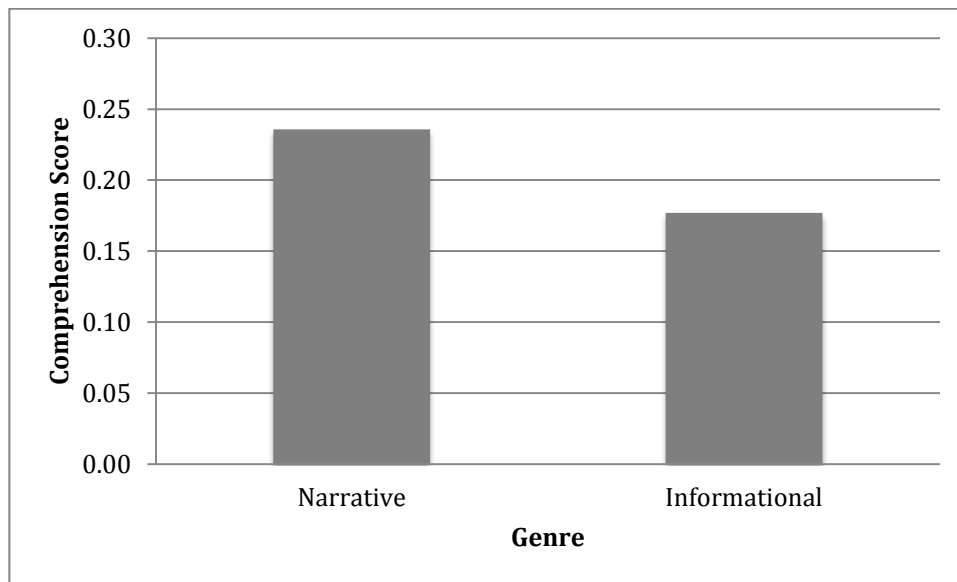


Figure 7. Main effect of genre on passage-specific reading comprehension.

Summary

This study evaluated the effects of two qualitative independent variables with two levels – genre (narrative/information) and referential cohesion (high/low) on oral reading fluency rate, oral reading fluency accuracy, and passage-specific comprehension. Results indicate that genre and referential cohesion have an interaction effect on rate and

accuracy, with strongest performance on the high cohesion narrative text. Performance on high cohesion narrative text was significantly greater than low cohesion narrative text and high cohesion informational text. Surprisingly, there was no effect of referential cohesion on passage-specific comprehension, on informational text reading accuracy, or on informational text reading rate. For passage-specific comprehension, there was a main effect for genre, indicating that students performed better on a measure of passage-specific comprehension on narrative texts than informational texts.

CHAPTER V

CONCLUSION

Discussion

The purpose of this study was to evaluate the effects of referential cohesion and passage genre on student reading proficiency (measured by oral reading fluency rate, accuracy, and passage-specific comprehension) within the context of curriculum-based measurement. The results of this study provide evidence that referential cohesion and genre affect student performance on oral reading fluency passages when readability is held constant. Specifically, these results indicate that high referential cohesion supports student rate and accuracy for narrative passages, but does not significantly increase oral reading fluency rate or accuracy for informational passages.

As outlined in the model of relations, it was hypothesized that genre and referential cohesion would have direct effects on oral reading fluency rate, accuracy, and passage-specific comprehension. As indicated in Figure 8, the study design allowed for evaluation of direct effects of genre and referential cohesion on the dependent variables. Results are consistent with hypothesized direct relations between: 1) referential cohesion and rate, 2) referential cohesion and accuracy, 3) genre and rate, 4) genre and accuracy, and 5) genre and comprehension.

Interpretations of non-significant relation between referential cohesion and comprehension. One potential interpretation of these findings is that referential cohesion may affect reading comprehension, but the selected measure failed to capture

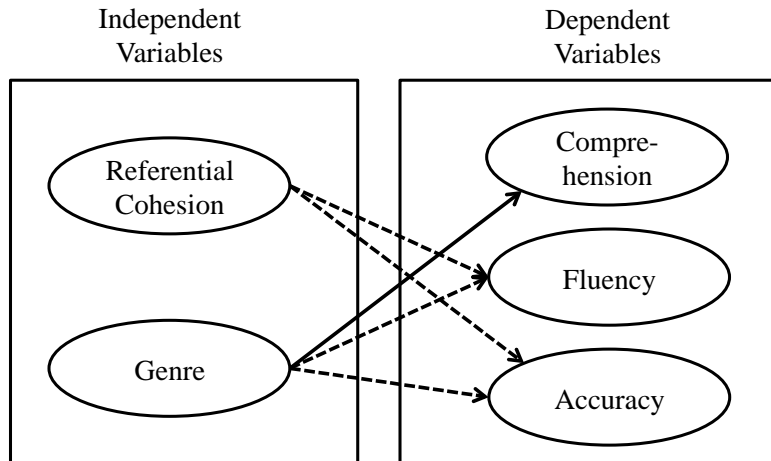


Figure 8. Revisited model of relations between independent and dependent variables. Dashed arrows represent interaction effects. Solid black arrow represents a direct, main effect.

these effects. While the recall task was selected only after careful consideration, it is possible that this task lacked the sensitivity to detect differences in comprehension performance. Reading comprehension is a large and complex construct, and existing technologies for measuring reading comprehension target individual features of understanding (such as the ability to retell a story using a high number of words or answer specific questions about events in the text). These challenges are articulated by Pearson & Hamm (2005):

Comprehension...is a phenomenon that can only be assessed, examined, or observed indirectly. We talk about the “click” of comprehension that propels a reader through a text, yet we never see it directly... We quiz them on “the text” in some way – requiring them to recall its gist or its major details, asking specific questions about its content and purpose, or insisting on an interpretation and

critique of its message. All of these tasks, however challenging or engaging they might be, are little more than the residue of the comprehension process itself. (p. 14)

This statement captures some of the challenges educators and researchers face in assessing reader comprehension of a text. As Pearson and Hamm (2005) argue, every measure of comprehension “carries with it a cost,” (p. 62) as researchers have yet to find a single measure that best captures the complex process of reading comprehension. For this study, a recall task was selected to measure the “residue” of the comprehension process; however, other measures may be used to measure passage-specific reading comprehension, such as multiple-choice or open-ended comprehension questions, cloze or maze procedures, or counts of words in a recall. It is possible that the selected comprehension measure was not sensitive to differences in passage-specific comprehension, or failed to capture the aspects of comprehension affected by referential cohesion; however, it is unknown whether other currently available alternatives would be any more sensitive. While it is possible that this measure was limited in sensitivity, it is also a strength of this design and was selected to provide the most sensitivity to effects possible, based on currently available technologies for measuring passage-specific comprehension.

A second potential interpretation of these findings is that referential cohesion only affects comprehension enough to increase fluency, but does not impact global understanding of the passage. For example, referential cohesion may reduce student hesitations or re-reads of text necessary for understanding by making connections between ideas explicit. While this would still impact rate of reading, it may not directly

impact comprehension, as the compensatory strategies named (hesitating or re-reading) allow for comprehension of the meaning of the idea unit. If students were limited to reading only the first minute of text, the impact on rate may more directly affect comprehension, as slower rate would limit the amount of text available to comprehend. However, because students in this study read the entire passage, differences in rate did not limit student comprehension score (i.e., students were exposed to the entire passage, even if rate was slow). In order to evaluate this hypothesis, future research should adjust scoring criteria to capture hesitations and repetitions and evaluate differences between high and low cohesion passages.

A third potential interpretation of these findings is that cohesion may impact comprehension, but that this effect is not stronger than individual and environmental contributors to reader comprehension. According to the members of the RAND Reading Study Group (2002), reader comprehension of text is based on three elements: the reader, the text itself, and the purpose for reading. The current study evaluated one feature (referential cohesion) of one of these elements (the text). While referential cohesion may have some effect on passage-specific comprehension, it is possible that this effect is overpowered by reader and environmental variables. For example, a reader might have a strong preference for narrative texts and consequently attend more to text meaning on narrative texts, which would overshadow any minimal benefits of referential cohesion.

Potential effects of background knowledge. One challenge in interpreting any measure of passage-specific comprehension is the potential effects of reader background knowledge. This is especially relevant to the present study, as students were only presented with one passage per condition. While background knowledge is a reader

variable and is challenging to control, it has the potential to have a strong effect on student performance on measures of comprehension. Additionally, there is evidence to suggest that background knowledge also contributes to oral reading fluency rate (Klauda & Guthrie, 2008). Consequently, educators must consider how to develop and build upon background knowledge in instruction, as well as identify how background knowledge may impact assessment tools.

Cohesion and grade level. The present study only evaluated the effects of referential cohesion for third grade students. While referential cohesion did appear to be important in the oral reading fluency of third grade students, it is possible that referential cohesion may impact performance differentially by grade level. For example, referential cohesion may have a stronger effect on reading performance in the early grades, because readers are still learning the alphabetic code and may rely on context clues to supplement limited decoding skills. Similarly, referential cohesion may be less important in later grades, as increased background knowledge and other student-level factors may have a greater impact on reading proficiency. Additional work is needed to understand whether effects can be generalized to the larger population of school-aged children.

Implications

Implications for instruction. The results of this study suggest that readers are impacted by the referential cohesion and genre of a passage, indicating a need for targeted instruction in approaching texts that may not inherently support fluent reading. Students may benefit from exposure to texts with high cohesion as well as low cohesion, so that students have strategies for decoding and understanding challenging texts when faced with when reading to learn and during assessment. As proposed in the Common

Core Standards (2010), educators should systematically introduce texts of higher complexity, including texts with lower referential cohesion. Students may benefit from instruction in explicitly identifying cohesive ties and using these ties to track passages meaning. Additionally, results indicate significant differences in student reading comprehension of narrative and informational texts, indicating that general comprehension instruction may be insufficient in supporting readers to comprehend various types of texts. Readers may benefit from comprehension instruction targeted to specific genres of texts; while general strategies may apply to all texts, students may need additional support in using effective strategies to extract information from informational texts. For example, students may benefit from instruction focused on purposes of reading (i.e., reading for enjoyment rather than reading for information), using structural elements in informational texts to identify information (e.g., table of contents, headings, tables and figures), and self-questioning strategies specific to informational texts (e.g., “What was this section about?” “What new words did I learn and what do they mean?”).

Implications for curriculum-based measurement. Analyses of student rate and accuracy scores indicate significant differences between passages due to referential cohesion and genre. However, it is necessary to consider the practical significance of these differences in interpretation of variability in CBM scores. Results of pairwise comparisons indicate significant differences in rate scores between the narrative/high referential cohesion passage and narrative/low referential cohesion passage (94.55 wcpm and 88.28 wcpm), as well as significant differences in rate scores between the narrative/high referential cohesion passage and the informational/high referential cohesion passage (94.55 wcpm and 90.97 wcpm). While these differences were statistically significant,

consideration of effect sizes indicates that these effects are very small. Additionally, it is necessary to consider how differences of this magnitude affect instructional practices.

Based on normative growth rates, a school may set a goal in which a student's rate score increases by two words per week. With such a goal, differences in rate scores from five to seven wcpm would affect educator interpretation of student oral reading fluency rate. However, the SEM for third grade for third grade DIBELS Next passages is reported as 11.29, indicating that the 95% confidence interval for a given rate score is +/- 22 wcpm (while the present study used non-standardized scoring procedures, the resulting scores were strongly correlated with standardized scores, so the SEM is likely still applicable). Therefore, differences in rate scores between passages of five to seven wcpm fall within the SEM of the third grade DIBELS Next passages. While such differences may have a meaningful impact on educator interpretation of student oral reading fluency skills, they are within the expected range for the selected passages.

The practical significance of differences in accuracy scores may be even more limited. Results of pairwise comparisons indicate significant differences in accuracy scores between the narrative/high referential cohesion passage and narrative/low referential cohesion passage (96% accuracy and 95% accuracy), as well as significant differences in accuracy scores between the narrative/high referential cohesion passage and the informational/high referential cohesion passage (96% accuracy and 95% accuracy). While these differences were statistically significant, effect sizes were small, indicating that differences were significant but minimal. Additionally, in practice there is little meaningful significance between 95% and 96% accuracy. Based on current research on oral reading fluency accuracy, both represent scores on texts that would be

considered at a student's instructional reading level (Hasbrouck, 1998). Consequently, differences in performance on oral reading fluency accuracy due to referential cohesion and genre will likely have little impact on educator interpretation of scores.

In addition to interaction effects on oral reading fluency rate and accuracy, the main effect of genre on comprehension may have an impact on assessment practices. Results of this study indicate a significant difference in passage-specific reading comprehension due to genre, with readers earning higher comprehension scores on narrative passages. The effect size of this difference was considered a medium effect ($d = 0.55$), indicating moderate practical significance. While existing CBM systems do not currently use the coded recall task for comprehension, many systems do integrate a measure of comprehension into the battery of benchmark measures (e.g., DIBELS Next Oral Reading Fluency-Recall and easyCBM Multiple Choice Reading Comprehension). Consequently, equivalency of alternate forms has implications for comprehension as well as oral reading fluency rate and accuracy. The results of the present study suggest that passage genre may impact equivalency of alternate forms in measuring comprehension. Specifically, these results indicate that students perform significantly better on narrative passages than informational. This poses a challenge for CBM development, as readers may comprehend and respond to narrative and informational texts differently. Consequently, test developers must consider means of reducing variability in student comprehension of narrative and informational passages. One approach, which expands upon the present study, is to continue to explore variables that may impact comprehension and include these variables in estimates of text complexity. A second approach is to administer both a narrative and an informational passage at each data

point, and compare performance within genres rather than between. Further research is necessary to evaluate the feasibility and practical benefits of these and other approaches to controlling for genre differences in reader comprehension.

Implications for measurement of text complexity. These findings suggest that genre and referential cohesion may contribute to the complexity of reading curriculum-based measurement oral reading fluency passages. Because Lexile (readability) scores were held constant across passages, these results indicate that Lexile scores did not entirely capture differences in passages due to genre and referential cohesion. While passage differences are inevitable in alternate forms, the results of this study indicate that such differences may have a meaningful impact on student performance on measures of oral reading fluency rate, accuracy, and comprehension. These findings provide evidence that referential cohesion and genre contribute to reading rate and accuracy; consequently, there may be a benefit in considering referential cohesion in estimates of passage complexity.

Additionally, this study presents a method of quantifying the effects of referential cohesion that is (a) feasible, and (b) sensitive enough to capture differences in passages. First, the use of a referential cohesion composite score to capture referential cohesion is easily accessible and feasible. The Coh-Metrix program is available to the public, and allows passages to be analyzed easily and quickly. The variables related to referential cohesion can be combined into a composite score using commercially available software, allowing for the measurement of the referential cohesion of selected passages. Second, these referential cohesion scores were capable of distinguishing between high and low cohesion passages, even within a set of passages that was designed to tightly control text

complexity. This is evidenced by differences in student performance as a function of this variable. Future research should further evaluate the validity of the RCCS in distinguishing between high and low cohesion passages by comparing the RCCS to expert ratings of the cohesiveness of selected passages.

Furthermore, these findings indicate that qualitative evaluation of text complexity may fail to fully capture the contribution of referential cohesion. Traditionally, the cohesiveness of a passage has been perceived as a qualitative feature of text, best evaluated through expert judgment and discussion (Common Core Standards, 2010). The text complexity of the selected passages was evaluated primarily by quantitative analyses (readability ratings); however, the authors of the measure report that anecdotal information was included in the overall assessment of passage difficulty. Based on the recommendations of the Common Core Standards, this qualitative analysis should be sufficient in capturing the variability due to text cohesion. However, the results of this study indicate that this qualitative analysis did not capture differences in the referential cohesion of the selected passages. As these differences were found to impact student reading performance, the ability to differentiate between highly cohesive and less cohesive passages appears to be an important feature of text complexity. Additional research is needed to determine if targeted qualitative analysis focused on referential cohesion can capture meaningful differences; however, these results suggest that current methods of qualitative analysis did not differentiate between the selected passages even though these differences were related to changes in student reading performance.

Study Limitations

One important limitation of the design is that each condition was represented by only a single passage. Consequently, it is difficult to know if the differences captured in analysis are due to condition or unique passage effects because the two are confounded. It is possible that there were characteristics of the selected passages that were not fully explained by Lexiles, genre, or referential cohesion estimates that impacted reading performance. Specifically, significant effects were associated with the high cohesion narrative passage; it is possible that there is something about that specific passage that aided in oral reading fluency rate and accuracy, in addition to or instead of high referential cohesion. It is also possible that readers responded to this specific passage differently than other passages, perhaps due to reader factors such as interest and background knowledge. Additional research should include additional passages for each condition in order to minimize passage effects due to text factors unrelated to referential cohesion.

A second limitation of this study is that it did not examine differential effects by skill level as an effect in the design. There is evidence to suggest that the effects of text cohesion on comprehension may vary based on reader proficiency (e.g., O'Reilly & McNamara, 2007). This design did not explore that issue and instead evaluated the effects of referential cohesion on readers across skill levels; however, it is possible that findings may not be applicable to subsets of students with very high or very low skills. Student skill level was not selected for inclusion in the study design because the focus of this work is text-based contributors to text complexity in the context of curriculum-based measurement (CBM). Because the same CBM passages are administered to all students

regardless of skill level, student skill level was not included as a central independent variable in this design. However, future research should evaluate the role of student skill level in text complexity both in CBM passages and other types of reading assessments.

A third limitation is that the administration procedures the DIBELS passages and scoring of oral reading fluency rate did not follow standardized procedures, which may affect the ability to generalize results to general outcome progress monitoring.

Correlations between the non-standardized rate score and first minute and easyCBM scores were strong, suggesting that the non-standardized, pro-rated procedure did not compromise the validity of the rate measure; however, follow up studies that more closely resemble standard CBM administration and scoring will be necessary to support the effects of genre and referential cohesion in educational practice.

A fourth limitation of this study design is that passages are controlled to only represent readability within a constrained range. It is possible that results will vary if readability scores were held constant at a lower or higher range (i.e., more or less readable texts). Consequently, future research should evaluate the role of referential cohesion on comprehension and reading rate for highly readable and less readable passages.

A fifth limitation is that passages were selected from a small sample of texts (third grade DIBELS Next Oral Reading Fluency passages), and effects may be sample-specific. Referential cohesion levels (high and low) were assigned based on ratings from the included set of passages; it is possible that other samples of passages (e.g., DORF passages from other grade levels or oral reading fluency passages from other curriculum-based measurement systems) may represent more or less variability in referential

cohesion scores. Future research should replicate the procedures for measuring referential cohesion used in this study with other sets of passages to evaluate the generalizability of these results.

A sixth limitation is the use of a coded recall task to measure comprehension. As discussed in Chapter 3, this measure may fail to capture meaningful differences in passage-specific comprehension. The use of different scoring schemes may yield different results, as might the use of different measures of passage-specific comprehension. Additionally, recall tasks rely on oral language skills, so it is possible that performance on the recall task was confounded with oral language ability. However, it is important to note that reading comprehension remains a difficult construct to measure across the field, and the selected measure was determined to be the best available tool to measure passage-specific reading comprehension.

A final limitation of the study design involves data collection and coding procedures. While inter-observer agreement data were collected on oral reading fluency scores, these data do not verify the procedural fidelity of administration of the oral reading fluency and passage-specific comprehension measures. The study would have been strengthened by the use of a procedural fidelity data collection tool, such as a checklist, to assess data collector fidelity to standardized data collection procedures. Additionally, all coding of student recalls was completed by a single researcher. Consequently, the reliability of the assigned codes is unknown. This may be remedied post-hoc by having a second researcher trained in the coding scheme verify coding of a proportion of recalls.

Next Steps

Replication. Replication allows for limitations to be addressed through small changes in study design. Perhaps most critical, replication is needed with more than one passage per study condition. As discussed above, it is possible that effects were due to differences in individual passages that were not captured by Lexile scores, genre, or referential cohesion measurement. Consequently, future work should include multiple passages in each condition.

In order to address possible differential effects by skill level, future work should include student skill level as an independent variable in the study design. Effects should be evaluated across the entire sample and by skill group in order to identify potential differences in performance.

Similarly, future research should expand the readability level of passages included in the study design. Not only did this study focus on a specific grade-level, but within that level passages were selected due to similarity of Lexile scores. Future work should replicate the basic study design using a greater range of passages, including passages and participants at various grade-levels, as well as passages within each grade-level with higher and lower Lexile scores. Future research should also replicate the study design using passages from a different source, to evaluate whether effects are isolated to the passages included in this study. Such research would allow for the generalization of effects beyond third grade DORF passages within a specified range of Lexile scores.

Future directions for measurement of referential cohesion. The results of this study suggest that it may be worthwhile to continue to explore means of measuring the referential cohesion of passages selected for assessment and instruction. This study

design presented one means of quantifying referential cohesion as a composite of the various devices within a text that support continuity of reference. Study results suggest that this composite may be sensitive to some differences in referential cohesion, possibly due to the inclusion of multiple devices that effect referential cohesion. It is recommended that future research continue to explore the use of a composite score, as individual devices may support referential cohesion while failing to capture the contributions of other devices. For example, two texts may represent similar levels of referential cohesion, but such cohesion may be accomplished in various ways. While one passage may maintain continuity of reference through the use of adjacent argument overlap, the second task may accomplish strong referential cohesion through a completely different device, adjacent anaphor overlap. The use of a composite score allows the referential cohesion of these passages to be compared, even though referential cohesion is maintained through different means. However, the measurement of individual devices may have value in understanding why a text has strong or weak referential cohesion. As in previous work on the effects of cohesive devices on reading comprehension (see Table 2), an examination of individual devices may allow educators and passage developers to revise texts to better support reader oral reading fluency rate and accuracy.

While the selected referential cohesion composite score demonstrates promise in the measurement of referential cohesion, future work should evaluate alternative means of creating a composite score to capture the referential cohesion of a passage. For example, future work should consider the weighting of the various cohesive devices in the creation of the composite score. For this research, all device scores were weighted

equally. However, the composite score may be strengthened if weighting of the individual devices were driven by a theory on relations between these devices. Additionally, future research should evaluate whether the inclusion of all devices in the composite score is necessary. In previous work, Hiebert (2011) created a referential cohesion composite score using only the argument overlap and stem overlap variables. Consequently, future work should evaluate whether the inclusion of additional devices contributes to the sensitivity to the composite score.

Finally, future work on the measurement of referential cohesion should consider the use of an external criterion for determining whether a passage represents high or low referential cohesion. In the present work, referential cohesion composite scores (RCCS) ranged from -1.12 to 3.17. However, it is unknown how this range of cohesion scores should be interpreted. Instead, this range of scores poses a number of questions that impact interpretation: Do these scores represent a wide or narrow range of referential cohesion? How do these scores compare to other methods of measuring referential cohesion? One means of beginning to address these questions is to compare quantitative estimates of referential cohesion to qualitative evaluation. For example, expert reviewers may assign referential cohesion ratings to passages, which can be compared to the RCCS. While this process alone would be insufficient to understand the range of the RCCS, it may help researchers understand if quantitative differences are detected and identified as meaningful through qualitative review.

Future directions in measurement of comprehension. It is recommended that future work continue to evaluate methods of measuring reading comprehension in relation to referential cohesion. One possibility is to explore secondary analysis of recall

data based on the comprehension codes assigned to recalled idea units. Rather than evaluating comprehension as a single score, future research may examine proportion or frequency scores for each type of recall response (conservative, liberal, no match-consistent, no match-inconsistent). In particular, future research should focus on differences in the no match-consistent score, as it is possible that highly cohesive texts support deeper comprehension, which may be captured by the extent to which a reader goes beyond what is stated explicitly in the text.

Additionally, future work should evaluate alternative methods of coding recalled responses. With the selected coding scheme, recalled idea units that were matched to an idea unit could only be assigned one of two codes: conservative and liberal. However, a qualitative examination of recalled idea units suggests variability in the types of responses within each code. In particular, the liberal code captured all responses that could be matched to an idea unit but did not represent a near verbatim recall of all relevant details. For example, responses for the original idea unit “Every day, Carrie and her teenage brother Jackson explored a new part of the preserve,” included: “every day they liked to go to a hike,” and “Carrie and her brother, Jackson, were going to take hikes at this preserve they found.” Both of these responses capture different components of the original idea unit – the first captures that the siblings hike every day, while the second omits “every day” but includes the detail that the siblings hike in the preserve. One alternative to the selected coding scheme is to assign a rating on an ordinal scale to each recalled idea unit based on alignment to the original idea unit.

Future work should also include an additional measure of comprehension in order to verify results using the coded recall. Consistent findings would provide support for the

use of a coded recall task as a means of measuring passage-specific comprehension with sensitivity. Inconsistent findings – specifically, significant relations between referential cohesion and reading comprehension – would indicate a need for further explanation of potential relations between referential cohesion and reading comprehension.

Summary

The complexity of text, which is defined by The Common Core Standards in English and Language Arts (2010) as the “inherent difficulty of reading and comprehending a text combined with consideration of reader variables (Glossary, p. 43),” has a number of implications for educators in the areas of instruction and assessment. Understanding and capturing the components that contribute to text complexity has implications for both instruction and assessment. Instructionally, the Common Core Standards Initiative (2010) stresses that students develop skills to be able to read and comprehend texts of increasing complexity as they progress through school. This expectation is based on data documenting the importance of comprehending complex texts in college and the workplace. In assessment, knowledge and understanding of text complexity has implications for both summative and formative assessment. For summative assessments such as state accountability tests, understanding of text complexity may help to improve test construction and interpretation. For formative assessment, controlling text complexity is critical in facilitating accurate individual decisions. Additionally, improved measures of text complexity will facilitate the development of better progress monitoring materials.

This study focused on the role of text complexity in assessment, specifically, formative assessment. Text complexity is particularly important in formative

assessments because such assessments utilize repeated, alternate, equivalent forms to capture student growth towards a general outcome or goal, and a key assumption of such tools is that alternate forms of the assessment are of equal complexity. Consequently, there is a need to better understand what variables contribute to text complexity, and how they impact student performance on formative assessments. This study was designed to evaluate features of text that are not typically included in readability estimates but may contribute to the complexity of the passage: passage genre and text cohesion.

Specifically, the study evaluated the role of text cohesion and genre on student oral reading fluency (reading with sufficient rate and accuracy) and comprehension performance, for the purpose of enhancing the utility and precision of formative assessment tools. Research questions addressed main effects for text cohesion and genre on reading rate, accuracy, and comprehension, and interactions between passage genre and text cohesion.

Univariate ANOVAs allowed for evaluation of direct effects of genre and referential cohesion on oral reading fluency rate, accuracy, and passage-specific comprehension. Results indicated effects for each of the dependent variables included in the study design. For oral reading fluency rate, results indicate a significant interaction between genre and referential cohesion on rate: when referential cohesion was high, rate was higher on narrative passages than informational passages. Follow up pairwise comparisons indicated that rate scores were significantly higher for the high cohesion narrative text than the low cohesion narrative text and the high cohesion informational text. For oral reading fluency accuracy, results also suggest a significant interaction between genre and referential cohesion on accuracy: high referential cohesion was related

to greater accuracy for narrative texts, while low referential cohesion was related to greater accuracy for informational texts. As with rate, pairwise comparisons indicated that accuracy scores were significantly higher for the high cohesion narrative text than the low cohesion narrative text and the high cohesion informational text. For passage-specific reading comprehension, there were no significant effects of referential cohesion. There was a significant main effect of genre on comprehension, with students performing significantly better on the passage-specific comprehension measure for narrative texts than informational texts. Altogether, these results indicate direct relations between both genre and referential cohesion on student reading performance.

The presence of these relations has implications for the development and interpretation of formative assessment tools. These findings indicate that genre and referential cohesion have a significant impact of student reading performance, and may contribute to complexity of reading CBM passages. Consequently, there is evidence that these features of text should be considered in estimates of text complexity. Additionally, this study provides evidence that referential cohesion may be able to be measured quantitatively. The metric used in this study, the Referential Cohesion Composite Score (RCCS), was easily developed using readily available technologies, and can be used to measure the referential cohesion of any set of passages. Results of this study indicate that the RCCS was able to differentiate passages with high and low referential cohesion, and that those differences were related to differences in oral reading fluency and accuracy.

REFERENCES CITED

- ACT, Inc. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*. Iowa City, IA: Author.
- Albano, A. D., & Rodriguez, M. C. (2012). Statistical equating with measures of oral reading fluency. *Journal of School Psychology, 50*, 43-59.
- Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. *Handbook of reading research, 1*, 255-291.
- Archer, A. L., Gleason, M. M., & Vachon, V. L. (2003). Decoding and fluency: Foundation skills for struggling older readers. *Learning Disability Quarterly, 26*, 89-101.
- Ardoin, S. P., Williams, J. C., Christ, T. J., Klubnik, C., & Wellborn, C. (2010). Examining readability estimates' predictions of students' oral reading rate: Spache, Lexile, and Forcast. *School Psychology Review, 39*, 277-285.
- Baker, D. L., Stoolmiller, M., Good, R. H., & Baker, S. K. (2011). Effect of reading comprehension on Passage Fluency in Spanish and English for second-grade English learners. *School Psychology Review, 40*, 331-351.
- Beck, I. L., McKeown, M. G., Omanson, R. C., & Pople, M. T. (1984). Improving the comprehensibility of stories: The effects of revisions that improve coherence. *Reading Research Quarterly, 19*, 263-277.
- Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman, J. A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly, 251-276*.
- Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing, 22*, 185-200.
- Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29*, 137-164.
- Best, R., Ozura, Y., Floyd, R. G., & McNamara, D. S. (2006,). Children's text comprehension: effects of genre, knowledge, and text cohesion. In *Proceedings of the 7th international conference on learning sciences* (pp. 37-42). International Society of the Learning Sciences.

- Briggs, R. N. (2011). Investigating variability in student performance on DIBELS oral reading fluency third grade progress monitoring probes: Possible contributing factors (unpublished doctoral dissertation). University of Oregon, Eugene, OR.
- Britton, B. K., & Gulgoz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology, 83*, 329-345.
- Cervetti, G. N., Bravo, M. A., Hiebert, E. H., Pearson, P. D., & Jaynes, C. A. (2009). Text genre and science content: Ease of reading, comprehension, and reader preference. *Reading Psychology, 30*, 487-511.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd Ed.)*. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cohen, S. A., & Steinberg, J. E. (1983). Effects of three types of vocabulary on readability of intermediate grade science textbooks: An application of Finn's transfer feature theory. *Reading Research Quarterly, 19*, 86-101.
- Common Core State Standards Initiative. (2010). *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.
- Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly, 42*, 475-493.
- Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin, 27*, 11-28.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American psychologist, 34*, 571-582.
- Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*, 184-192.
- Deno, S. L., & Marston, D. (2006). Curriculum-based measurement of oral reading: An indicator of growth in fluency. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about fluency instruction* (pp. 179-203). Newark, DE: International Reading Association.
- Duran, N. D., Bellissens, C. Taylor, R. S., & McNamara, D. S. (2007). *Quantifying text difficulty with automated indices of cohesion and semantics*. Proceedings of the 29th Annual Meeting of the Cognitive Science Society. Austin, TX: Cognitive Science Society.

- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9, 167-188.
- Elfenbein, A. (2011). Research in text and the uses of coh-metrix. *Educational Researcher*, 5, 246-248.
- Florida Center for Reading Research (2006). Empowering teachers. Retrieved from <http://www.fcrr.org/assessment/et/resources/glossary3.html>
- Foorman, B. R. (2009). Text difficulty in reading assessment. In E. H. Hiebert (Ed.), *Reading more, reading better* (pp. 231-250). New York, NY: Guilford.
- Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology*, 46, 315-342.
- Freebody, P., & Anderson, R. C. (1983). Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. *Reading Research Quarterly*, 18, 277-294.
- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5, 239-256.
- Good, R. H., Kaminski, R. A., Cummings, K., Dufour-Martel, C., Petersen, K., Powell-Smith, K., Stollar, S., & Wallin, J. (2011). *DIBELS next*. Eugene, OR: Dynamic Measurement Group.
- Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K. A., & Latimer, R. J. (2011). *DIBELS next technical manual*. Eugene, OR: Dynamic Measurement Group.
- Graesser, A.C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3(2), 371-398.
- Graesser, A. C., McNamara, D., S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40, 223-234.
- Graesser, A. C., McNamara, D., S., Louwrese, M. M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36, 193-202.
- Graves, M. F., & Graves, B. B. (2003). *Scaffolding reading experiences: Designs for student success* (2nd ed.). Norwood, MA: Christopher-Gordon.

- Greenfield, G. (1999). *Classic readability formulas in an EFL context: Are they valid for Japanese speakers?* Unpublished doctoral dissertation, Temple University, Philadelphia, PA.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London, England: Longman.
- Hasbrouck, J. E. (1998). *Reading fluency: Principles for instruction and progress monitoring*. Professional Development Guide. Austin, TX: Texas Center for Reading and Language Arts, University of Texas at Austin.
- Hiebert, E. H. (1998). *Text matters in learning to read* (Report 1-001). Ann Arbor, MI: Center for the Improvement of Early Reading Achievement.
- Hiebert, E.H. (2001). Standards, assessment, and text difficulty. In A.E. Farstrup & S.J. Samuels (Eds.) *What research has to say about reading instruction (3rd Ed.)*. Newark, DE: International Reading Association.
- Hiebert, E. H. (2011). *Using multiple sources of information in establishing test complexity* (Reading Research Report 11.03). Santa Cruz, CA: TextProject, Inc.
- Hiebert, E. H., & Fisher, C. W. (2007). Critical word factor in texts for beginning readers. *Journal of Educational Research, 101*, 3-11.
- Hiebert, E.H., & Pearson, P.D. (2010). *An examination of current text difficulty indices with early reading texts* (Reading Research Report 10.01). Santa Cruz, CA: TextProject, Inc.
- Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003a). Accuracy and fluency in list and context reading of skilled and RD groups: Absolute and relative performance levels. *Learning Disabilities Research and Practice, 18*, 237-245.
- Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003b). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*, 719-729.
- Jungjohann, K. (2010, January). *Reading and writing in the content area*. Lecture conducted at the University of Oregon, Eugene, OR.
- Kame'enui, E. J., & Simmons, D. C. (2001). Introduction to this special Issue: The DNA of reading fluency. *Scientific Studies of Reading, 5*, 203-210.
- Kendeou, P., & van den Broek, P. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology, 97*, 235-245.

- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.
- Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly*, 10, 62-102.
- Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology*, 100(2), 310-321.
- Ledoux, K., Traxler, M. J., & Swaab, T. Y. (2007). Syntactic priming in comprehension: evidence from event-related potentials. *Psychological Science*, 18(2), 135-143.
- Lennon, C., & Burdick, H., (2004). *The Lexile framework as an approach for reading measurement and success* [white paper]. Retrieved from <http://www.learningwithjamesgentry.com/Resources/TAKSDP/Lexile-Reading-Measurement-and-Success-0504.pdf>
- Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, 94, 778-784.
- Lynch, J. S., & van den Broek, P. (2007). Understanding the glue of narrative structure: Children's on-and off-line inferences about characters' goals. *Cognitive Development*, 22, 323-340.
- Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*, 21, 251-238.
- McDonald, J. H. (2009). *Handbook of biological statistics* (2nd ed.). Sparky House Publishing: Baltimore, MD.
- McKeown, M. G., Beck, I. L., Sinatra, G. M., & Loxterman, J. A. (1992). The contribution of prior knowledge and coherent text to comprehension. *Reading Research Quarterly*, 27, 78-93.
- McMaster, K. L., van den Broek, P., Espin, C. A., White, M. J., Rapp, D. N., Kendeou, P., Bohn-Gettler, C. M., & Carlson, S. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences*, 22, 100-111.
- McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, 55, 51-62.

- McNamara, D. S., Graesser, A. C., Cai, Z., & Kulikowich, J. M. (2011). *Coh-metrix easability components: Aligning text difficulty with theories of text comprehension*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22*, 247-288.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.
- McNamara, D.S., Louwrese, M.M., Cai, Z., & Graesser, A. (2005). Coh-Metrix version 1.4. Retrieved from <http://cohmetrix.memphis.edu>.
- McNamara, D. S., Louwrese, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47*, 292-330.
- MetaMetrix (2013). *Lexile-to-grade correspondence*. Retrieved from <http://www.lexile.com/about-lexile/grade-equivalent/grade-equivalent-chart/>
- Morris, J., & Hirst, G., (1991). Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics, 17*, 21-48.
- National Assessment Governing Board (2008). *Reading Framework for the 2009 National Assessment of Educational Progress*. Washington, DC: American Institutes of Research.
- O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes, 43*, 121-152.
- Ozuru, Y., Briner, S., Best, R., & McNamara, D. S. (2010). Contributions of self-explanation to comprehension or high- and low-cohesion texts. *Discourse Processes, 47*, 641-667.
- Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction, 19*, 228-242.
- Parker, R. I., Vannest, K. J., Davis, J. L., & Clemens, N. H. (2010). Defensible progress monitoring data for medium- and high-stakes decisions. *Journal of Special Education, XX(X)*, 1-11.

- Pearson, P. D. (1974). The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. *Reading Research Quarterly*, 155-192.
- Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices – Past, present, and future. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment*. Mahwah, NJ: Lawrence Erlbaum.
- Pikulski, J.J., & Chard, D.J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, 58(6), 510–519.
- Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 669-682). New York, NY: Academic Press.
- Powell-Smith, K. A., Good, R. H., & Atkins, T. (2010). *DIBELS next oral reading fluency readability study* (Technical Report No. 7). Eugene, OR: Dynamic Measurement Group.
- RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Risko, V. J., & Walker-Dalhouse, D. (2011). Drawing on text features for reading comprehension and composing. *The Reading Teacher*, 64, 376-378.
- Roberts, R., Good, R., & Corcoran, S. (2005). Story retell: A fluency-based indicator of reading comprehension. *School Psychology Quarterly*, 20, 304-317.
- Saenz, L. M., & Fuchs, L. S. (2002). Examining the reading difficulty of secondary students with learning disabilities: Expository versus narrative text. *Remedial and Special Education*, 23, 31-41.
- Stage, S. A., & Jacobsen, M. D., (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*, 30, 407-419.
- Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16, 32-71.
- Stanovich, K. E., & West, R. F. (1981). The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 658-672.
- Tapiero, I. (2007). *Situation models and levels of coherence: Toward a definition of comprehension*. Mahwah, NJ: Lawrence Erlbaum Associates.

- van den Broek, P., & Gustafson, M. (1999). Comprehension and memory for texts: Three generations of reading research. In Goldman, S. R., Graesser, A. C., & van den Broek, P. (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 15-34). Mahwah, NJ: Lawrence Erlbaum Associates.
- Vidal-Abarca, E., Martinez, G., & Gilabert, R. (2000). Two procedures to improve instructional text: Effects on memory and learning. *Journal of Educational Psychology, 92*, 107-116.
- Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment, 11*, 85-104.
- Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 386-397.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162-185.
- Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading, 2*, 199-220.