



Minnesota State University, Mankato
Cornerstone: A Collection of Scholarly
and Creative Works for Minnesota
State University, Mankato

All Graduate Theses, Dissertations, and Other
Capstone Projects

Graduate Theses, Dissertations, and Other
Capstone Projects

2022

The Ability of DIBELS to Predict Proficiency on the Minnesota Comprehensive Assessment and the WIDA ACCESS for ELLs

Katherine Schultz
Minnesota State University, Mankato

Follow this and additional works at: <https://cornerstone.lib.mnsu.edu/etds>



Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Language and Literacy Education Commons](#)

Recommended Citation

Schultz, K. (2022). The ability of DIBELS to predict proficiency on the Minnesota Comprehensive Assessment and the WIDA ACCESS for ELLs [Master's thesis, Minnesota State University, Mankato]. Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato. <https://cornerstone.lib.mnsu.edu/etds/1247/>

This Dissertation is brought to you for free and open access by the Graduate Theses, Dissertations, and Other Capstone Projects at Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato. It has been accepted for inclusion in All Graduate Theses, Dissertations, and Other Capstone Projects by an authorized administrator of Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato.

**The Ability of DIBELS to Predict Proficiency on the Minnesota Comprehensive
Assessment and the WIDA ACCESS for ELLs**

Katherine Schultz

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctorate in Educational Leadership at Minnesota State University, Mankato

Mankato, Minnesota

May, 2022

April 1, 2022

The Ability of DIBELS to Predict Proficiency on the Minnesota Comprehensive Assessment and the WIDA ACCESS for ELLs

Katherine Schultz

This dissertation has been examined and approved by the following members of the student's committee.

Dr. Natalie Rasmussen
Advisor

Dr. Nancy Drescher
Committee Member

Dr. Paul Peterson
Committee Member

Table of Contents

ABSTRACT	IV
CHAPTER 1.....	1
INTRODUCTION	1
<i>Background of the Problem</i>	<i>1</i>
<i>Problem Statement</i>	<i>5</i>
<i>Research Questions.....</i>	<i>6</i>
<i>Significance of the Research</i>	<i>6</i>
<i>Delimitations (and Limitations)</i>	<i>7</i>
<i>Definition of Terms</i>	<i>8</i>
CHAPTER 2.....	14
REVIEW OF LITERATURE.....	14
<i>Theoretical Framework of Oral Reading Fluency</i>	<i>15</i>
<i>Recent Educational Reform in America</i>	<i>18</i>
<i>Accountability Measures and English Language Learners.....</i>	<i>21</i>
ACCOUNTABILITY TESTING IN MINNESOTA.....	23
<i>Minnesota Comprehensive Assessments (MCAs).....</i>	<i>23</i>
<i>WIDA for English Language Learners.....</i>	<i>25</i>
PROGRESS MONITORING USING READING FLUENCY	29
<i>Defining Oral Reading Fluency</i>	<i>29</i>
<i>Reading Fluency and English Language Learners</i>	<i>31</i>
CBM-R AND DIBELS.....	32
<i>Curriculum-Based Measures in Reading (CBM-r)</i>	<i>33</i>
<i>Dynamic Indicator of Basic Early Literacy Skills (DIBELS).....</i>	<i>35</i>
<i>DIBELS and ELLs.....</i>	<i>38</i>
CHAPTER 3.....	42
RESEARCH QUESTIONS AND NULL HYPOTHESES	42
RESEARCH DESIGN	44
PARTICIPANTS	44
INSTRUMENTS.....	45
CHAPTER 4.....	48
SAMPLE AND METRICS	48
DATA ANALYSIS.....	50
<i>MCA and DIBELS.....</i>	<i>50</i>
SUMMARY	59
LIMITATIONS	59
CHAPTER 5.....	61
INTRODUCTION	61
SUMMARY OF THE RESEARCH.....	61
SUMMARY OF THE FINDINGS	63
IMPLICATIONS	65
RECOMMENDATIONS	67
REFERENCES.....	73

Abstract

The Ability of DIBELS to Predict Proficiency on the Minnesota Comprehensive Assessment and the WIDA ACCESS for ELLs

Katherine Schultz

Doctorate in Educational Leadership

Minnesota State University, Mankato, Minnesota

2022

Recent educational reform gave birth to accountability testing and development of student sub-groups meant to track gaps in achievement. Because of the high-stakes nature of these assessments, educators have sought efficient progress monitoring tools. Oral reading fluency's link to overall reading success has made it a desirable means of measuring growth. This link is well researched with native English speakers but less researched with English learners. This research seeks to determine if the commercially-produced oral reading fluency suite of assessments, DIBELS, can predict reading success on the Minnesota Comprehensive Assessments (MCAs) in English learners to the same degree as with native English speakers. Additionally, the research sought to find if a correlation could be made between the MCAs and the WIDA ACCESS test (Minnesota's language proficiency assessment). Archival student data from a rural Minnesota school district was analyzed using Pearson Product Moment Correlation to determine if a correlation exists between the DIBELS suite of assessments, the MCA in Reading, and the WIDA ACCESS. In addition, regressions were used to determine the predictive power of DIBELS for native English speakers and English learners. Results showed the correlation between DIBELS and MCA was greater for native English speakers than for English learners. It was also more predictive of reading success for native English speakers than for English learners. Also, the ACCESS reading subtest showed the strongest correlation to the MCA in reading. Overall, until English learners reach proficiency in English, their assessment results are less valid than for native English speakers.

Chapter 1

Introduction

Background of the Problem

Over the past three decades, the demographic makeup of our school populations has changed sharply. According to the Forum on Child and Family Stats (2019), 50% of the students registered in American schools are children of color, Hispanic, Asian, and other linguistically and culturally diverse groups. It is predicted that these numbers will continue to skyrocket over the next 30 years with the students of color representing over 60% of the schools. Because of this growing trend and the diverse linguistic and cultural backgrounds, these students will bring to their schools, districts, and teachers must be prepared to sufficiently address their unique needs. This becomes particularly essential when assessing these unique groups. Although the idea of developing assessments that consider more inclusive universal designs has been practiced since the passage of the Individuals with Disabilities Education Act of 2004 (Public Law No. 108-446), researchers have found that these design considerations do not adequately address the needs of all populations, including English Language Learners (ELLs) (Liu & Anderson, 2008). Although modifications and accommodations are often employed in an attempt to diminish inequalities between native English-speaking students and non-native speakers, they do not balance the two. Because ELLs are not yet proficient in the language they encounter on their state-mandated assessments, their scores are often seen as a deficit and the results are misused in determining their academic ability and performance capabilities (Mahoney, 2008). These misuses can lead to higher dropout rates, grade retention, and

higher rates of needing to complete the General Education Diploma (GED) versus a traditional high school diploma (Amrein & Berliner, 2003).

At the district and school level, ELLs can also be disadvantaged when performing on summative assessments, especially in reading. English Language Learners (ELLs) possess the unique ability to decode or “sound out” words within their new or second language. “Cracking the code” using phonics knowledge has led many ELLs to become “word callers” (Quirk & Beam, 2012). Their reading approximates native-like fluency but lacks native-like comprehension. This causes difficulty when assessing an ELL’s true reading ability.

An assessment commonly used to gain a quick snapshot of students’ reading abilities is a Curriculum-Based Measure or CBM (Baker & Good, 1995). This assessment involves a one-minute timing of students reading orally (Wiley & Deno, 2005). The overwhelming evidence in the research suggests these timings can accurately predict overall reading comprehension and future reading success in native English speakers (Cutting & Scarborough, 2006; Fuchs et al., 2001; Muyskens et al., 2009). In fact, Good et al. (2001) found that native English-speaking students reaching benchmark on oral reading fluency measures such as CBMs were likely to meet or exceed expectations on high-stakes, standardized tests. Although there are studies that show CBMs to be an effective measure when working with language minority students (Baker & Good, 1995; Betts et al., 2008; Yesil-Dagli, 2010), some caution that ELLs may be overlooked or misidentified when oral reading fluency data is the only piece of information to indicate a need for additional reading support (Quirk & Beem, 2012).

A form of CBM that has gained popularity since the advent of the CBM is DIBELS (Dynamic Indicator of Basic Early Literacy Skills). Founded as a not-for-profit through the College of Education at the University of Oregon, the DIBELS suite of assessments can be found in thousands of schools across this country as well as internationally (University of Oregon Center on Teaching and Learning, 2021). DIBELS has proved popular because of its ease of delivery and low cost. Additionally, it is sought after for its usefulness as a progress-monitoring tool as it can be administered frequently throughout the course of the school year (Good et al., 2001). Although the creators of DIBELS argue its subtests allow teachers to predict underachievement in reading and thus recommend appropriate interventions and monitor student growth, the research is not as positive in regard to ELLs. Scheffel, Lefly, and Houser (2012) found no significant studies that sufficiently tested the reliability of DIBELS as a means to specifically assess English language learners. They went on to state that few rigorous studies have been conducted to demonstrate the predictive ability of the DIBELS assessments in determining which children will need additional intensive interventions to achieve reading success. Muyskens et al. (2009) also state that many school professionals question the validity of the use of these assessments with ELLs as many are able to read fluently yet comprehend at a much lower level.

Numerous studies show a strong correlation between DIBELS Oral Reading Fluency (ORF) scores and proficiency on state-mandated standardized assessments (Buck & Torgeson, 2003; Shaw & Shaw, 2002; Vandermeer et al., 2005). However, studies that specifically pinpoint its effectiveness in predicting success in ELLs are less

certain. Scheffel et al. (2012) showed that DIBELS was most effective at identifying ELLs who were potentially “at-risk” of not meeting proficiency on standardized tests yet more accurately identifying non-ELLs who are “low risk.” Kim et al. (2016) found that the predictive ability of DIBELS ORF scores varied based on students’ language levels. This adds an additional element of uncertainty in the research when considering how ELLs move through the language learning levels.

In Minnesota, the English proficiency level is measured using the World-class Instruction and Design in Assessment (WIDA) Assessing Comprehension and Communication in English State-to-State (ACCESS) assessment (WIDA, 2021). Each year, ELLs are required to demonstrate their proficiency in listening, speaking, reading, and writing the English language. The assessment is built on English Proficiency Standards that have been linked to the Minnesota Academic Standards (Chi & Lin, 2012). Based on their results, students are given a proficiency level ranging from Level 1 to Level 6. Although it would seem as though Level 6 would indicate native-like proficiency, the state of Minnesota has defined proficiency as reaching an overall composite score of 4.5 with three out of the four domains (listening, speaking, reading, and writing) reaching at least 3.5 (Minnesota Department of Education Student Support Division, 2017). Additional research is needed to determine what role DIBELS ORF scores play in predicting academic success in English Language Learners and how their language learning level as indicated on the WIDA ACCESS test may impact those results.

Both English learners and Native English Speakers in Minnesota must also demonstrate overall academic proficiency on a state-wide standards-based assessment known as the Minnesota Comprehensive Assessment (MCA) in reading and math. This assessment begins in third grade and continues each year through eighth grade. To reach proficiency students must have a score of the student's grade level multiplied by 100 plus 50 (Minnesota Department of Education, n.d.). For example, a third-grader would need to achieve a minimum score of 350 (third-grade x 100 + 50). A fourth grader would need to achieve a minimum score of 450 (fourth grade x 100 + 50). The MCA in reading and math is given in English only with a minimal amount of translation allowed when giving test directions.

Problem Statement

Often, English learners are subjected to scrutiny equal to that of their native-speaking peers. They are given the same assessments and required to perform to the same degree as those who are native to the language of the test. This can often lead to misconceptions about the true abilities of these students. Invalid inferences can be made about their skills and this leads one to question the construct validity of the test (Sireci et al., 2008). Sireci et al. (2008) refer to this as construct-irrelevant variance with English learners. Attempts have been made to combat this variance. Various accommodations have been proposed and allowed to try to minimize the impact language has on measuring the construct being tested. Not only is this variance affecting English learners on high-stakes tests but its impact can also be felt at the district, school, and individual levels (Bailey & Carroll, 2015). Districts can be impacted by funding related to the

success of these students. Initiatives and programs can be falsely validated or invalidated based on the results of standardized tests (Wolf et al., 2008). Schools can be rewarded or penalized for the perceived success or failure of these tests. Individually, students can be affected by their scores. Decisions about placement in classes, eligibility for certain offerings that might interest students, and access to core content are all made based on results of tests that could have validity concerns (Bailey & Carroll, 2015). At the classroom level, teachers often make decisions about placement in intervention programs based on data from summative assessments (Goffreda et al., 2009).

Research Questions

1. Can DIBELS ORF scores accurately predict proficiency on the Minnesota Comprehensive Assessment (MCA) for English Language Learners in grades 3 and 4?
2. Does proficiency on the DIBELS ORF and MCA lead to proficiency on the WIDA ACCESS test?

Significance of the Research

English learners' cognitive abilities are often questioned when their language proficiency affects their performance on standardized tests. They are tested in a language for which they are not yet proficient but the results are analyzed equally against students whose native language is English. This research addresses this bias. Often, ELLs who show proficiency on DIBELS may be overlooked for reading concerns as they appear to be proficient. Yet, scores on the MCA show they are not yet comprehending at the necessary level. Interventions could have been put into place had other assessment pieces

been considered. Conversely, students can be over-identified for reading interventions when their DIBELS scores are significantly lower than their English-speaking peers. Rather than focusing on language development, teachers and interventionists may choose to focus on phonemic and phonics skills. Again, this could translate into gains on DIBELS but not show growth in comprehension. With this research, districts could begin to progress monitor and assess English language learners in more effective ways. Looking at skills other than reading fluency could point educators toward an applicable intervention or program for these unique learners.

Delimitations (and Limitations)

In this research, English language learners are compared to all mainstream students in grades 3 and 4. Within the “all” category, there is no distinction between special education students and non-special education students. Not all special education students participate in the alternative test (Minnesota Tests of Academic Skills). Many are required to take the MCA and their data is included in the findings of this research. That may be significant as their learning disabilities most certainly affect their performance on these tests. Their disabilities could also impact their performance on DIBELS assessments as well. It was decided to leave them in the “all” data set. Future research would address these unique learners.

In the English learner group, no distinction has been made between language and cultural groups. Multiple languages are represented in this group of students. Most students are of African descent but Latinx, Asian, and Middle Eastern students are also represented in the data set. No allowances are made for these linguistic differences. For

example, Latinx students whose alphabetic characters are the same as English may have an advantage over Asian and Middle Eastern students whose alphabetic characters are significantly different from English characters. This linguistic distance may impact the ease with which students can acquire reading skills. Additionally, no distinction is made between students who are literate in their first language and those who have not acquired literacy skills in their first language. First language literacy generally makes second language literacy significantly easier. Unfortunately, “cracking the code” in one’s first language can also lead to “word calling” in a second language. Students can read phonetically thus scoring higher on oral reading fluency assessments without having an equal ability to understand the words being read.

Definition of Terms

Accuracy. The percentage of words read correctly. On the DIBELS assessment, accuracy is calculated by dividing the number of words read correctly by the total number of words read (<https://dibels.uoregon.edu/training/bir/accuracy-and-fluency.php>). It is recommended by DIBELS that students demonstrate an accuracy score of 97% or higher for an independent reading level.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS). DIBELS is a collection of short, one-minute reading assessments. These assessments are designed to allow educators to universally screen students, diagnose early reading difficulties, and progress monitor growth toward grade-level benchmarks (<https://dibels.uoregon.edu/assessment/dibels/dibels-eighth-edition>). The DIBELS suite of assessments includes five subtests, targeting specific reading skills:

- Letter Naming Fluency (LNF) - Students must quickly and accurately name alphabetic letters, both uppercase and lowercase. The frequency with which each letter appears in the assessment is a reflection of their frequency in real-world text.
- Phoneme Segmentation Fluency (PSF) - The assessment identifies individual sounds within words. Students are given one minute to divide each given word into its individual phonemes. The beginning stages of phonological awareness are identifying the initial sound, moving to the middle sound, and finally distinguishing the final sounds as well (Yesil-Dagli, 2011).
- Nonsense Word Fluency (NWF) - Students must demonstrate their knowledge of letter-sound correspondences by decoding unfamiliar words (Scheffel, et al., 2012). Students can choose to identify all the sounds in the word and/or recode those sounds to produce the entire word. Proctors record the total number of sounds correctly identified and the number of whole words read or recoded.
- Word Reading Fluency (WRF) - This is the newest assessment added to the DIBELS suite. Students are given a list of grade-appropriate words that they need to read with automaticity. If a student hesitates for more than three seconds on any word, the word is provided to the student and marked as incorrect.
- Oral Reading Fluency (ORF) - This one-minute timing evaluates students' ability to read grade-level connected text fluently and accurately. The number of words read correctly is compared to the grade-level benchmarks established by the

DIBELS organization to determine if students are reading at a fluent enough level.

English Language Learners. In Minnesota, an English language learner must indicate they speak a language other than English on the Minnesota Language Survey (MNLS) (Minnesota Department of Education, 2017). If a student identifies they learned a language other than English first, they must be screened using an approved English Language Proficiency (ELP) screener. Since Minnesota employs the WIDA suite of assessments when assessing students, the WIDA Screener is used to determine eligibility for English language services.

Minnesota Comprehensive Assessment (MCA) and the Minnesota Test of Academic Skills (MTAS). To meet all state and federal requirements, students must complete the MCA or MTAS to show proficiency or growth toward Minnesota's academic standards (<https://education.mn.gov/MDE/fam/tests/mca/>). Mainstream students in grades 3-8 and grade 10 take the MCA in reading and students in grades 3-8 and grade 11 take the MCA in math. Students who meet specific requirements receiving special education services can take the alternate MTAS.

Oral Reading Fluency. Oral reading fluency is often defined in different ways by different scholars. Most agree that oral reading fluency involves the ability to blend sounds into meaningful words automatically and efficiently (Baker et al., 2011; Fuchs et al., 2009; Wise et al., 2010). The differences lie in more specific elements. Some argue that reading fluency encompasses vocabulary knowledge (Yesli-Dagli, 2011). Others argue that personal experiences and inferring also lead the reader to more accurately read

and comprehend the text (Wise et al., 2010). This research will focus primarily on automaticity and speed while reading.

Proficiency. In this research, proficiency will be used in two different ways. First, when referencing the MCA/MTAS, proficiency is shown by achieving a score of the student's grade level multiplied by 100 plus 50 (Minnesota Department of Education, n.d.). For example, a third-grader would need to achieve a minimum score of 350 (third-grade x 100 + 50) to achieve proficiency with the Minnesota academic standards. Proficiency when considering English language learners is determined quite differently.

In Minnesota, ELLs must also take the WIDA ACCESS test to determine their language proficiency in English. WIDA has established six levels of proficiency starting with level 1 (students with very low proficiency) and ending with level 6 (native-like proficiency). To demonstrate proficiency, students must reach a composite score of 4.5 to be considered proficient enough to succeed in mainstream instruction. Additionally, students cannot score below 3.5 in any one of the four tested areas of listening, speaking, reading, and writing (Minnesota Department of Education, 2017).

Progress Monitoring. Because of the high-stakes nature of standardized testing, monitoring growth throughout the academic year has become vital. Using assessments that model the summative standardized tests becomes impractical. Delivering these large tests frequently throughout the school year eats away the instructional time and promotes testing fatigue in students. For that reason, educators have adopted quick measures such

as DIBELS to monitor student progress. One-minute timings of students' reading is far more practical and efficient than longer, more cumbersome tests.

Second Language Acquisition. Second Language Acquisition (SLA) is a branch of Applied Linguistics that seeks to understand how learners acquire a second or non-native language subconsciously (Aljumah, F.H., 2020; Krashen, 1982). Acquiring a second language is distinguished from learning a second language in the research. Students can learn language rules and patterns but that does not necessarily lead to fluent use of the target language. The goal should be acquisition (Krashen, 1981).

World-class Instruction Design and Assessment (WIDA). Developed by the University of Wisconsin and the Wisconsin Department of Education, the WIDA suite of assessments targets English language learners and the academic language required for them to succeed in an English-speaking school. Minnesota has joined the consortium and uses the WIDA assessments to evaluate the proficiency of their English learners. There are two main tests: The WIDA screener and the WIDA ACCESS (Assessing Comprehension and Communication in English State-to-State).

The WIDA screener is used to qualify students for English language support. Eligibility is first determined by the Minnesota Language Survey. If a language other than English is indicated on the survey, students must be screened with the WIDA screener (Minnesota Department of Education, 2017). The screener tests academic language in the areas of listening to, speaking, reading, and writing English. Students who score a composite below 4.5 on the screener qualify for English language support from a Language Instruction Education Program (LIEP).

WIDA ACCESS. The ACCESS is an annual assessment that English learners currently enrolled in a Language Instruction Education Program (LIEP) must take to demonstrate English proficiency. Like the WIDA screener, the ACCESS involves four separate assessments that measure proficiency in the four language domains identified by WIDA: listening, reading, speaking, and writing. To meet proficiency standards, English learners must achieve an overall composite score of 4.5 or higher with no domain dipping lower than 3.5 (WIDA, 2020).

Chapter 2

Review of Literature

Attempts at educational reform in America over the past 30 years have brought high-stakes testing into the spotlight. How to best measure academic progress and growth has divided lawmakers and educators. Lawmakers require evidence of growth while educators know that growth does not always show itself in standardized ways, especially when considering the diversity of today's student body. English language learners must not only learn a new language but must also demonstrate the same academic proficiency as native English speakers. In Minnesota, this growth is measured on the Minnesota Comprehensive Assessment test given annually. English language learners must also show language proficiency on a separate language test also given annually.

Given the high-stakes nature of these assessments, how can one progress monitor effectively to ensure academic success on yearly exams? Oral reading fluency has become the key skill in determining future academic success. Most agree that oral reading fluency is directly linked to reading comprehension and is a valid predictor of future reading achievement. Yet, the research is far less conclusive when considering the uniqueness of English language learners functioning in English-speaking schools. Can oral reading fluency effectively predict future reading success in English learners as well as it can for native English speakers?

Theoretical Framework of Oral Reading Fluency

How children best learn to read has been in constant debate and evolution since the early 1960s. The instructional pendulum steadily swings between bottom-up theory and top-down theory as curriculum developers and instructional coaches try to determine what best leads to successful readers.

Proponents of the Bottom-up theory start with the smallest unit of meaning (phonemes) to the largest (understanding and comprehension) (Gregory, 2016). They see reading as combinations of letters and sounds to produce words that produce meaning. Many see this as the “phonics approach” to reading (Gregory, 2016). Behaviorism, a theory focusing on changing behavior over time, supports a bottom-up approach. It breaks the complex task of reading into smaller sub-skills with direct instruction coming from the experienced person (the teacher) to the less experienced person (the student) (Aldhanhani & Abu-Ayyash, 2020). The teacher provides the instruction and delivers crucial feedback to the learner to build their capacity to read fluently. This theory has received harsh criticism as it sees students as passive learners who simply receive instruction/feedback and produce the correct output (Aldhanhani & Abu-Ayyash, 2020). The information processing model more accurately demonstrates how students’ brains develop language and reading fluency. This theory stresses the importance of learning to read fluently so the reader can utilize their energy and memory for comprehending the text versus expending it on decoding the text (Aldhanhani & Abu-Ayyash, 2020). Again, critics claim this theory depicts students as computers simply absorbing information to produce the desired outcome. The

Automaticity Theory also depends on the speed at which word recognition occurs (LaBarge & Samuels, 1974). The more attention students must give to decoding, the less attention they can provide to the task of understanding the text (Aldhanhani & Abu-Ayyash, 2020). It encourages a “drill and kill” methodology, and critics have faulted it as too time-consuming and impractical in the classroom. Finally, the Ehri Word Learning Model theory is another widely recognized theory that focuses on bottom-up thinking about reading. This theory has five stages readers go through as they begin to build their oral reading fluency skills (Metsala & Ehri, 1998). Each of these five stages deals with the most fundamental components of decoding, starting with preschool skills and moving towards more advanced levels that develop speed and automaticity (Metsala & Ehri, 1998). Schools that utilize fluency-based measures such as the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are more likely to focus on bottom-up reading skills such as phonemic awareness and phonics skills (Good III et al., 2002).

An early cognitive psychologist, Ausubel (1968) distinguished between meaningful learning and rote learning, leading the way for the development of top-down reading approaches and theories. Top-down theorists contest that meaning develops first from previous knowledge and experiences. The reader is merely responsible for predicting written words based on what makes sense using their existing knowledge of semantic and syntactic rules (Gregory, 2016). Chang et al. (2020) demonstrated that oral language proficiency does indeed impact reading success and significantly impacts the effectiveness of reading instruction, supporting the ideology of the top-down approach. Smith (1994) argued that reading is not a passive, mechanical process as bottom-up

theorists would posit but rather purposeful, rational, and dependent on the experiences the readers bring to the text. Nunan (1999) further developed this thinking with the “schema theory.” This theory focuses on how past experiences help to frame our understanding of what we are currently experiencing. Anderson (1994) presented research related to how a reader’s schemata impact how they comprehend the messages presented in the text. If the content of the text is familiar, reading the text becomes less difficult. Pre-reading, during reading, and post-reading activities help the reader develop schemas they may not already have and/or activate the schema they do have (Villanueva de Debat, 2006). Others argue that top-down reading strategies such as Repeated Readings are more effective in developing reading comprehension (Taguchi et al., 2012). When using the repeated readings model in combination with the auditory model, learners listen to and reread passages several times, allowing them to learn pronunciations of new words, guess at word meanings, and understand at the word, phrase, sentence, and passage level more effectively (Taguchi et al., 2012).

When considering second language acquisition, top-down theories and strategies are applied most often when thinking about literacy and oral reading fluency. One of the more well-known theorists in second language acquisition, Stephen Krashen, hypothesized in his Monitor Theory that learning grammatical and structural rules leads to “learning” a language when the goal should be acquisition (Krashen, 1981). When language learning occurs through linguistic rules and patterns, students monitor their language usage too much, and their language output can be affected or hampered by this overuse (Krashen, 1981). Krashen’s theory also included the Comprehensible Input

Hypothesis, which he later amended to the Compelling Comprehensible Input Theory. Students are more likely to acquire vocabulary and grammatical structures when the input they receive is compelling and comprehensible (Krashen & Bland, 2014; Ng et al., 2019). Additionally, extensive reading for pleasure and specialized learning leads to more remarkable language development and reading comprehension development (Krashen & Bland, 2014; Ng et al, 2019). Friesen and Haigh (2018) hypothesized that teaching reading strategies versus rules and grammar allow second language learners the flexibility to acquire more language as they encounter it in text.

Considering top-down reading theories are often recommended by second language acquisition theorists, utilizing bottom-up theories and assessments when assessing the reading ability and growth of language learners seems counterintuitive. Can bottom-up assessments such as DIBELS accurately demonstrate the reading ability of English learners, and additionally predict the future reading success of these unique learners? Does the development of bottom-up reading skills in an effort to reach benchmark scores on assessments such as DIBELS lead to greater comprehension skills in English learners to the same degree as native English speakers?

Recent Educational Reform in America

In 1983, Secretary of Education T.H. Bell released findings from his newly created National Commission on Excellence in Education on the state of education in America to the federal government. This document, known as *A Nation at Risk*, spearheaded decades of educational reform policies and agendas in the United States (National Commission on Excellence in Education, 1983). Further, The National

Commission on Excellence in Education assessed the quality of education provided through public and private schools at all levels, including colleges and universities across the country. Additionally, the commission compared educational outcomes in the United States to results in other advanced nations worldwide. Their findings were grim as they announced, without doubt, that the educational system in America had been stripped down to an unconscionable level. At that time, the commission found that approximately 23 million Americans were functionally illiterate, including 17-year-olds who made up 13% of that number. More startling, the number of illiterate minority students could have been as high as 40%:

Part of what is at risk is the promise first made on this continent: All, regardless of race or class or economic status, are entitled to a fair chance and to the tools for developing their individual powers of mind and spirit to the utmost. (National Commission on Excellence in Education, 1983, *The Risk*, para. 5)

The report delivered recommendations to bolster the educational systems currently in place: laying out the number and types of courses required for high school graduation, focusing attention on basic literacy skills, and encouraging colleges and universities to ramp up admittance policies and procedures.

Educational reform and policy change became a priority as each incoming president adopted and changed what the previous commander in chief accomplished, or failed to achieve, in the areas of curriculum and standards development and assessment (Hayes & Williams, 2008). No Child Left Behind, the most notable reform, came as a reauthorization of the Elementary and Secondary Education Act (ESEA) of 1965, which

was meant to address the needs of minority students by providing additional support and funding to schools in disadvantaged areas (Bunch, 2011). This law required states to develop rigorous academic standards, create assessments to address these standards adequately, and monitor school performance annually, with the goal of 100 percent proficiency by 2014 (Bunch, 2011). Not only would states need to monitor growth on their assessments each year in grades 3 - 12, but they must also monitor efforts to close the gap between White students and students of color and economically advantaged versus disadvantaged students. Five ethnic categories were developed: White, Black, Hispanic, Asian or Pacific Islander, and American Indian. Three additional categories were created to encompass students qualifying for free or reduced-price lunches, English Language Learners, and those receiving special education services (Hayes & Williams, 2008). Schools were to show progress towards their goals and ultimately 100 percent proficiency by 2014 or face harsh and more harsh penalties each year that targets were missed (Hayes & Williams, 2008).

Positive outcomes were witnessed over the next decade, albeit small gains. Black students had shown growth in Vocabulary and Math on the Scholastic Aptitude Test (SAT); dropout rates decreased in most populations (except for Latinos); and high schools toughened their attendance policies, thereby increasing the number of students in classes each day (Walters, 1993). Many schools also reported lengthened school days and the adoption of “no pass, no play” rules in their athletic departments (Walters, 1993). Teacher salaries also increased by 22 percent in that decade (Walters, 1993). Yet, the gap was still evident.

The consequences of the accountability era in education have been felt for years, well beyond the decade following No Child Left Behind and its subsequent reauthorizations. The annual state-created assessment has become the only meaningful measure of success. It has encouraged eliminating any subject matter not tested, with reading, and math being at the center of all instruction (Neill, 2003). Because of the grave consequences associated with failing scores, schools and districts have resorted to teaching to the test, proving proficiency has become the objective of the classroom rather than measuring teaching and learning (Smyth, 2008). Consequently, teachers have been hesitant to innovate in the classroom or experiment with diverse strategies to increase learning for all students because of the uncertainty of their impact on test results (Smyth, 2008).

Accountability Measures and English Language Learners

Although No Child Left Behind (NCLB) brought often forgotten student populations into the spotlight, assessing English language learners (ELLs) has proven difficult. In the NCLB law, Congress required states receiving federal Title funds to develop assessments that tested five different content areas: comprehension, listening, speaking, reading, and writing (Bunch, 2011). The urgency to create meaningful assessments is confounded by the increased representation of English learners in our schools. By 2050, roughly 40% of the total student population will be comprised of language learners (Goldenberg, 2008). School achievement in the U.S. will be highly dependent on the test scores of these unique students (Lazarin, 2006). Although attempts have been made to create programs and assessments that meet language learners' needs,

ethnically diverse students continue to fail at a much greater rate than the white, non-Hispanic population (Reyes & Rorrer, 2001). Even though strides have been made to utilize more universal design considerations when developing standardized tests, these steps cannot adequately address the bias present in “one-size-fits-all” assessments (Liu & Anderson, 2008).

Not only do English learners encounter cultural bias in their assessments, but they are also judged unfairly by the structure of the current test reporting. Students counted as ELL have been identified as limited in English and are currently participating in an English development program (Ardasheva et al., 2012). Students who have completed an English program and have exited that program are not considered in score reporting, creating the perception that ELLs are intellectually less capable than their white, non-Hispanic peers (Aradasheva et al., 2012). Additionally, standardized tests do not evaluate academic ability in students’ first language. They rely solely on the language in which the students are not yet proficient (Reyes & Rorrer, 2001). In *The Standards for Educational and Psychological Testing*, The American Educational Research Association published its concerns about the validity of testing when language proficiency is limited,

Any test that employs language is, in part, a measure of their language skills...This is of particular concern for test takers whose first language is not the language of the test... In such instances, test results may not reflect accurately the qualities and competencies intended to be measured. (AERA et al., 1999, p. 91)

Unfair comparison is then made between those limited in English and those who are native to the language, even when test accommodations are employed to attempt to increase the validity of the assessment (Sireci et al., 2008). Construct validity is challenged and becomes irrelevant when content areas are assessed using English with English learners (Mahoney, 2008). Not only does that assessment represent a content area, but it also challenges the language ability of English language learners, hence becoming a language assessment as well (Sireci et al., 2008). Regardless of validity concerns, states are required to assess ELLs equally with non-ELLs using accommodations and modifications to further limit the bias encountered by the learners (Reyes & Rorrer, 2001).

Accountability Testing in Minnesota

Minnesota Comprehensive Assessments (MCAs)

In Minnesota, prior to the reauthorization of the Elementary and Secondary Education Act, students in third, fifth, eighth, and eleventh grades were targeted for yearly accountability testing. Now, all students in grades three through eight must participate in reading and math assessments each year, along with tenth grade for reading and eleventh grade for math (Welsh, 2003; Minnesota Department of Education, n.d.). Under the World's Best Workforce statute (Minnesota State Statute 120B.11, 2020), schools must work to:

meet school readiness goals; have all third-grade students achieve grade-level literacy; close the academic achievement gap among all racial and ethnic groups of students and between students living in poverty and students not living in

poverty; have all students attain career and college readiness before graduating from high school; and have all students graduate from high school. (p. 1)

The Minnesota Comprehensive Assessments (MCAs) and the Minnesota Test of Academic Skills (MTAS) are meant to measure student progress toward meeting Minnesota's academic standards and goals (Minnesota Department of Education, n.d.). Not only is its purpose measuring student progress towards meeting Minnesota's academic standards, but it also meets the federal and state legislative requirements for yearly accountability measurement (Minnesota Department of Education, n.d.). Beginning in 2015, Minnesota adopted Computer Adaptive Tests (CAT), eliminating the need for traditional, static assessments of the past (Deatz, et al., 2013). Utilizing this format allows the assessment to adapt and change based on the answers chosen by the test-taker. Alignment studies conducted by independent researchers have shown this format best demonstrates student ability and increases the validity of the results shown (Deatz, et al, 2013). Missing from this research is mention of the validity with English Language Learners.

Upon completing their yearly test, all students are given a scale score and a designation that corresponds to that score. A benchmark score of the students' grade level $\times 100 + 50$ equals "Meets" proficiency. For example, a third-grader would need to achieve a score of 350 ($3 \times 100 + 50$), a fourth-grader would need a score of 450 ($4 \times 100 + 50$), and a fifth-grader would need a score of 550 ($5 \times 100 + 50$). Any student scoring below but within 10 points of that benchmark would earn the designation "Partially Meets." Any student scoring ten points or more above that benchmark would be

designated “Exceeds” (Minnesota Department of Education, n.d.). Systems are judged based on the number of students meeting or exceeding that benchmark, with partial credit awarded for students designated as “Partially Meets.”

Beyond academic achievement and progress, accountability in Minnesota also includes progress toward English language proficiency, graduation rates, and consistent attendance (Minnesota Department of Education, n.d.). Additionally, student scores are evaluated across sub-group categories: English language learners, students qualifying for free or reduced lunch, special education students, and students falling into seven different ethnic or racial categories (Minnesota Department of Education, n.d.). Yearly goals are set for each district and school to meet, with consequences attached to continuously failing schools (Welsh, 2001). Consequences for failing schools can include targeted support or comprehensive support based on their scores. Targeted support is required when one or more of the sub-groups identified previously fall below the lowest five percent of all Title I schools (Minnesota Department of Education, n.d.). Comprehensive support is much more extensive and occurs when a school’s score falls below the lowest five percent of all Title I schools. A customized plan is developed based on that school’s context, needs, and student population (Minnesota Department of Education, n.d.).

WIDA for English Language Learners

In addition to the MCAs in reading and math, ELLs must also take an English Language Proficiency (ELP) Assessment that determines their proficiency in listening to, speaking, reading, and writing in English (Minnesota Department of Education, n.d.). Along with 39 other states, Minnesota has adopted the WIDA (World-Class

Instructional Design and Assessment) suite of assessments to screen potential students and assess current English learners (WIDA, 2020). WIDA was developed first by the Wisconsin Department of Public Education and later moved to the University of Wisconsin-Madison. Not only were assessments developed to assess language learners across content areas such as math, science, and language arts (Bunch, 2011), English Language Proficiency Standards were created to guide participating states in delivering appropriate instruction and rigor (WIDA, 2020). Rather than focusing on language learners' perceived deficits, WIDA adopted a "Can Do Philosophy" focusing on the positive attributes and skills these learners bring into their learning environment at each language proficiency level.

When enrolling in a Minnesota school, families must fill out the Minnesota Language Survey (MNLS). This document asks four questions regarding which language the student first learned or currently uses most frequently at home (Minnesota Department of Education, 2017). Following an MNLS that identifies the student as having a primary language other than English, an age-appropriate screener must be utilized, focusing on language proficiency across content areas. This screener is meant to demonstrate a student's ability to access grade-level content successfully. The WIDA screener is an online assessment that identifies students in need of further language development. Students who indicate a language other than English on the MNLS are next required to complete the WIDA screener in Minnesota (Minnesota Department of Education, 2017). It is also useful in determining their current language proficiency level, which allows instructors to more accurately plan and deliver appropriate instruction

(WIDA, 2020). Students in grades 1 - 12 complete the screener in an online format that includes 4 subtests in reading, listening, speaking, and writing. Students who do not show proficiency will be offered English language development instruction (Minnesota Department of Education, 2017).

Once a student has been identified as ELL, they begin taking the WIDA Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS) test yearly to monitor or demonstrate English proficiency. Like the MCA, the ACCESS test is given online and has been developed to adjust questions based on students' responses (WIDA, 2020). Again, the ACCESS tests language proficiency in listening, speaking, reading, and writing in English. Scores are divided into six proficiency levels: Entering (1), Emerging (2), Developing (3), Expanding (4), Bridging (5), and Reaching (6). A student receives a score in each tested area, with an overall composite score derived from the four individual scores (WIDA, 2020). Each state in the consortium determines exit criteria using ACCESS scores (Bunch, 2011). Over the past 5 years, Minnesota has adjusted its exit criteria to more accurately coincide with MCA proficiency scores (Minnesota Department of Education, 2017). Currently, ELLs in Minnesota must show an overall composite score of 4.5, with at least three out of the four domains (listening, speaking, reading, or writing) showing 3.5 or higher.

In Minnesota, the Department of Education limits the amount of time students can spend testing each school year to ten hours, including state and district assessments combined (Minnesota Department of Education, n.d.). If a student were to enter a

Minnesota school and have an MNLS that indicates a language other than English, they would be screened using the ACCESS screener, which takes roughly 80 minutes to complete (WIDA, 2020). Additionally, they would also take the MCA in math and reading (and possibly science if 5th or 8th grade), which could take up to 80 minutes for each test. Finally, they would take the WIDA ACCESS, which takes approximately 265 minutes to complete all four domains (WIDA, 2020). Add in any additional district-level assessments, and they could potentially meet or exceed that limit quickly. Unquestionably, testing fatigue could also impact the validity of the scores.

The question most often raised by educators and researchers is: Can these high-stakes tests accurately measure and predict academic performance, especially when considering the unique needs of English language learners. Is there a correlation between success on the MCA and success on the WIDA ACCESS? If there is a strong correlation between each of these assessments, how can one monitor progress towards these benchmarks which only occur annually? Because of these assessments' high-stakes nature and the unique nature of second language learners, early signs of reading difficulties must be caught and addressed as soon as possible (Kim et al., 2016). Districts need to monitor the progress of their students regularly and closely to ensure benchmarks are met at the year's end (Ostayan, 2016). Giving students MCA-like assessments at intervals throughout the school year is impractical and leads to significant instructional time loss. Quick, easy, progress monitoring assessments are necessary to save this essential resource (Kim et al., 2016).

Progress Monitoring Using Reading Fluency

Defining Oral Reading Fluency

The publication of the National Reading Panel's recommendations for reading instruction in 2001 brought reading fluency into the educational spotlight as an indicator of future reading and overall academic success. Yet, the focus on reading fluency began long before its emphasis in this publication. LaBerge and Samuels (1974) first drew attention to reading fluency by proposing their Automaticity Theory which highlighted the necessity for students to free their attentional capacity from lower-level skills such as decoding and word identification to allow this attention to focus on higher-level skills such as comprehension. The National Reading Panel (2001) agreed with this theory stating that fluency will enable students to focus their mental energies on interpreting and comprehending texts. Few researchers contest this idea, and reading fluency has become the focus of most reading progress monitoring tools and assessments (Goffreda et al., 2009; Kim et al., 2016; Vanderwood et al., 2014).

Although defining reading fluency seems simple on the surface, many have dug into the complexity of what true reading fluency encompasses. Baker et al. (2011) broadly define oral reading fluency to include: "vocabulary knowledge, lexical access, semantic skills, syntactic understanding, background knowledge, and literal and inferential comprehension" (p. 332). Fuchs et al. (2001) generalize reading fluency as complex skills that competent readers perform effortlessly, thus allowing the evaluation of reading fluency also to determine overall reading competence. Wise, et al. (2010) more explicitly defines reading fluency to encompass the automatic correspondence

between letters and sounds, the blending of those sounds into recognizable lexical chunks, forming links both within and between sentences, making connections between the text, and prior knowledge, and inferring missing information. Kuhn et al. (2010) not only consider accuracy and automaticity in their definition of oral reading fluency but also include prosody's role in reading fluency. They believe that proper use of prosodic features when reading can offer evidence of how the reader understands the text.

Regardless of the specificity and breadth of the definition offered, few argue the link between oral reading fluency and overall reading ability; that as fluency increases, comprehension follows suit. The National Reading Panel (2001) stated that fluency allows students to focus their cognitive energies on interpreting and comprehending text. Pikulski and Chard (2005) consider fluency to be the link between decoding and comprehension. Burns et al. (2002) believe fluency to be the minimum reading rate necessary for comprehension to occur. Other studies also consider vocabulary knowledge and factors such as prosody that may affect the reader's ability to comprehend (Kuhn et al., 2010). Yovanoff et al. (2005) not only attribute reading success to reading fluency but also include high-level vocabulary knowledge. They propose that as students progress through the grade levels, their comprehension will improve as their rate of reading and vocabulary knowledge develops. Based on their study, they maintain that reading fluency and vocabulary development can explain 40% to 50% of the variance in reading comprehension.

Reading Fluency and English Language Learners

The link between oral reading fluency and comprehension becomes more complicated when considering the uniqueness of English language learners (ELLs). Kim et al. (2012) recognize the difficulty encountered by English learners when faced with the complex task of reading in a language in which they are not yet fluent. In their 2012 study, Quirk and Beem provided evidence of a similar general relation between reading fluency and comprehension in ELL students and previous studies performed with non-ELL populations. They took a close look at students identified by teachers and administrators as “word callers.” This included students capable of decoding words fluently but who lacked sufficient comprehension skills. Their study showed that 55.5% of the total sample had reading fluency scores that were slightly higher than their reading comprehension scores. They also discovered that between grades 2 and 5, students’ fluency rates might be increasing at a faster rate than their reading comprehension. With ELLs, they found that as students became more proficient in English, their reading fluency and comprehension increased as well. Yet, they also found the highest level of word callers at the intermediate level.

Other studies have shown that oral language proficiency in ELLs may play a significant role in reading fluency. Kim (2012) hypothesized that oral language fluency would influence reading fluency rates and that students with developed oral language fluency would outperform those with limited oral language fluency. Indeed, the study found that participating students’ oral language skills in their second language (L2) were directly related to their L2 reading comprehension regardless of their reading rate. In

fact, the child's oral language skill was not associated with their oral or silent reading fluency. The author suggests the students' oral language proficiency may need to reach a certain level before it can uniquely contribute to their L2 reading fluency. Crosson and Lesaux (2010) also hypothesized that underdeveloped L2 oral language may impact the relationship between reading fluency and comprehension. Their study did find that reading fluency and comprehension are affected by L2 oral language, which raises the question regarding the effectiveness of measuring oral reading fluency in language minority students. Even though there are concerns about their effectiveness with all students, especially ELLs, and factors such as listening comprehension and vocabulary development can affect their sensitivity, oral reading fluency is often relied on for monitoring student progress at the classroom level.

CBM-r and DIBELS

Oral reading fluency measures as a means of progress monitoring have become a common tool for measuring reading growth (Valencia, et al., 2010). These assessments are considered sensitive enough to detect small changes in students' academic development and can be administered frequently (Fewster & Macmillan, 2002). Because of their sensitivity to change, oral reading fluency assessments can be used in decision-making, such as grouping instructional groups, determining the need for reading interventions, setting goals, and monitoring academic progress (Yovanoff et al., 2005). Not only can these measures be used in decision-making but can also be used to evaluate the effectiveness of instruction both within the classroom and in specific reading interventions (Schilling et al., 2007). Yet, some argue that measures of oral reading

fluency measures that focus only on rate and accuracy lead to misconceptions of the construct of reading, which in turn leads to instructional practices focusing only on tasks and skills that will lead to increased assessment results (Kuhn et al., 2010). Valencia et al. (2010) recommend expanding oral reading fluency measures to include other characteristics of the reading construct to ensure results are an accurate depiction of the student's ability. Regardless, oral reading fluency assessments continue to be the most popular. Two such measures, Curriculum-Based Measures in Reading (CBM-r) and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), are the most commonly employed.

Curriculum-Based Measures in Reading (CBM-r)

CBM-r are one-minute timings of students' reading as the teacher listens and marks errors. The students are then given a score of words correct per minute (wcpm) (Valencia et al., 2010). The one-minute reading passages are chosen from the students' curriculum allowing teachers to gain insight into each student's understanding of the instruction that has been delivered. According to Fuchs and Deno (1991), CBM-r contain three essential elements: test materials are drawn directly from the classroom curriculum, the tests are repeated multiple times over the course of an extended amount of time, and the results of the assessment are used to make instructional decisions for each student. CBM-r are attractive to educators as they are efficient and straightforward to administer, the results are easily interpreted, and they are inexpensive (Valencia et al., 2010). Fewster and MacMillan (2002) point out that the efficiency in delivering this

assessment allows for the frequent administration necessary to measure growth in reading.

Although research points out many positive attributes of using CMB-r when measuring reading growth, it also mentions negative qualities. Because reading passages are chosen from the classroom curriculum, a large amount of time is required on the instructor's part to locate appropriate passages to use. Also, the difficulty level can vary both within the text and between texts (Kuhn et al., 2010). There can also be variations based on whether students are involved in literature-based reading programs or skill-based reading programs (Hintz & Shapiro, 1997). Fewster and MacMillan (2002) caution that CBM-r was not created to be used in isolation but rather to be used in conjunction with other assessments. It was meant to be used as a general indicator of reading success or failure. Valencia et al. (2010) express concern about the narrow nature of CBM-r. This assessment measures only rate and accuracy. Other elements of the construct, such as phrasing, expression, and comprehension, are not directly measured. Of even more significant concern, Ardoin, et al. (2013) worry the results about the validity and reliability of CBM-r are often over-generalized and used for purposes for which there is insufficient empirical data. The authors continue by saying that progress-monitoring outcomes lack reliability and validity unless they are collected over an extended period of time. Although there are significant concerns about the use of CBM-r, they are still preferred over the traditional standardized test for showing student growth because of their sensitivity to small change and their ability to impact instruction (Kuhn et al., 2010).

Although the research is a bit limited in relation to the use of CBM-r with ELLs, there are studies that show a positive correlation between them. Muyskens et al. (2009) found the use of CBM-r in the fall was an accurate predictor of the students' spring reading assessment scores. The authors showed that three out of four ELLs could be classified on their spring assessment based on their fall CBM-r score. They continue to recommend the use of CBM-r with ELLs when making decisions for problem-solving models such as Response to Intervention (RTI). Baker and Good (1995) compared the CBM-r wcpm scores of English-only students and bilingual students. They found no significant difference in the CBM-r scores between the English-only and bilingual students. They both scored roughly the same number of words read correctly at the beginning of the study and throughout the study. Yet, when given comprehension assessments, the English-only students scored higher than their bilingual peers. Teachers also rated the English-only students as better overall readers than the bilingual students. From the study, the authors concluded that CBM-r was a reliable assessment of reading ability, including ELLs' reading ability. They also concluded that CBM-r are a reliable tool for monitoring the reading growth of ELLs. Because the reading passages are drawn directly from the curriculum which is being taught, ELLs can perform more equally with their English-only peers.

Dynamic Indicator of Basic Early Literacy Skills (DIBELS)

Because of the variability in individual teacher choices and curriculum types, generic, commercial versions of the CBM-r have been developed, the most commonly known version being DIBELS, or Dynamic Indicator of Basic Early Literacy Skills

(Kuhn et al., 2010). According to Manzo (2005), DIBELS has become a catchphrase both in schools and within state departments as a means of informing instruction, identifying students who may be at risk for future reading failure, and holding schools accountable. The DIBELS suite of assessments covers a range of developmental tasks and includes benchmarks to determine whether students are reading at grade level or are at risk for future reading failure (Manzo, 2005). The tasks assessed by DIBELS include reading components that researchers have identified as essential in early elementary reading (Schilling et al., 2007). These tasks include Letter Naming Fluency, Phoneme Segmentation, Initial Sound Fluency, Nonsense Word Fluency, Word Reading Fluency, and Oral Reading Fluency (<https://dibels.uoregon.edu/resources>). In each of these subtests, students are allowed one minute to produce as many items as possible with the goal of reaching a benchmark score three times a year (Ostayan, 2016). DIBELS also provides progress monitoring options to allow teachers to assess students as often as deemed necessary (Ostayan, 2016). This ability to progress monitor allows teachers to track students who may be showing signs of potential reading difficulties (Godffreda et al., 2009).

Although DIBELS is used extensively throughout the country, many have concerns about its use or overuse, even among native English speakers. Manzo (2005) points out educators' tendency to teach to the test or give the test too much weight when considering reading ability. Schilling et al. (2007) caution that DIBELS should not be the only instrument used when assessing students' true literacy skills. Numerous studies have been conducted regarding the effectiveness of the DIBELS suite of assessments in

predicting future reading success. In their 2007 study, Schilling et al. found that certain subtests of the DIBELS suite that test foundational and developmental reading skills (Initial Sound Fluency, Nonsense Word Fluency, Phoneme Segmentation Fluency), lost their predictive ability as students moved through first grade by the end of the year, only the Oral Reading Fluency (ORF) subtest was effective at predicting reading success. They went on to test the reliability of the ORF subtest in predicting success on summative reading assessments, such as the MCA found in Minnesota. They found that students identified as “some risk” and “low risk” of reading failure by the DIBELS ORF assessment were not properly identified. Seventy-two percent of second graders who fell in the “some risk” category and thirty-two percent of the second graders in the “low risk” category fell short of meeting the benchmark on the end of the year reading test. However, they found that students identified as “at-risk” were more accurately identified for potential reading failure. Eighty percent of second graders and seventy-six percent of third-graders in the “at-risk” category scored below the 25th percentile on their spring assessment. The authors recommend that teachers use a combination of both “at-risk” and “some risk” at the Fall ORF to identify students who could fall below the fiftieth percentile on their spring reading assessment. Still, they caution the use of DIBELS for making valid and reliable decisions about future student performance.

Good et al. (2009) also tested the DIBELS suite of assessments; however, they more specifically tested the benchmark levels set by the DIBELS creators. They suggested that the first-grade benchmark goal of 40 words correct per minute (wcpm) on the ORF was the most significant predictor of future reading success. Of the students

who reached the first-grade benchmark of 40 wcpm, 97% also reached the second-grade benchmark goal. The kindergarten suite of assessments was less likely to predict future reading success, with only 55% of students benchmarking in kindergarten continuing to benchmark in first grade.

DIBELS and ELLs

Teachers and school practitioners have raised concerns concerning the use of oral reading fluency measures in decision making when considering students who decode words fluently without an equal rate of comprehension, as is the case with English language learners (Muyskens, et al., 2009). Quirk and Beem (2012) refer to this group of students as “word callers” and also include ELLs in this category. Muyskens et al. (2009) verify the difficulty in evaluating and predicting students’ reading ability whose first language is not English. The authors state it is not uncommon for school employees to question the validity of oral reading fluency measures with ELLs because of their ability to decode words for which they may not have corresponding background knowledge or contextual experience. They also raise concerns about the limited amount of research available on the use of oral reading fluency measures with ELLs. They state that even though the research on CBM-r is extensive, it is not as thorough when applied to the needs of ELLs.

Because of the growing popularity of generic probes of oral reading fluency such as the DIBELS suite of assessments, concern has been raised as to their reliability when considering students who may read words faster than they comprehend words, specifically ELLs. Reliability is crucial considering oral reading fluency probes are often

the only measure used when making decisions about educational interventions (Quirk & Beem, 2012). Yet, according to Scheffel et al. (2012), no studies could be found that thoroughly tested the reliability and validity of the DIBELS suite of assessments on English Language Learners. Quirk and Beem (2012) caution that the use of oral reading fluency probes such as DIBELS will cause an overestimation of the reading comprehension skills of a large number of ELLs.

In their 2010 study, Valencia et al. cautioned relying solely on wcpm measures such as DIBELS when making decisions about students' overall reading ability. They examined student data in two ways. They first looked at "false negatives." This included students who were categorized by wcpm as low risk or some risk for reading failure who in fact failed to meet benchmark expectations on reading comprehension assessments. Based on their wcpm score, these students would not be identified as needing any sort of reading intervention (50th – 74th percentile), yet eventually fell below the 25th percentile on a comprehension assessment. The authors also approached their study from the vantage point of the reliability and sensitivity of the measure being used. They stated that an assessment that perfectly identifies students at risk for failure would be considered 100% sensitive. According to Johnson, Jenkins et al. (2009), an acceptable level of sensitivity for an assessment would fall somewhere between 90% and 95% sensitive. When testing the sensitivity of the DIBELS suite of assessments, Valencia et al. (2010) found the sensitivity to be at 77% for 2nd grade, 78% for 6th grade, and 78% for 4th grade. They found that across all grade levels, five of the six grade levels demonstrated inadequate levels of sensitivity. The authors recommend that multiple

assessments be used to accurately identify students at risk for reading failure, especially when considering populations who may already have difficulty with reading comprehension.

Although Valencia et al. (2010) found the sensitivity of the DIBELS suite of assessments to lack appropriate sensitivity, Scheffel et al. (2012) found the sensitivity of these assessments to be higher with ELL students than with non-ELL students. They found that between 51% and 64% of students classified as “at-risk” by the DIBELS assessments were correctly classified whereas 92% to 93% of students classified as “low risk” were correctly classified. This indicated that the DIBELS assessments more accurately predicted reading success than reading failure. Surprisingly, the percentage of ELL students accurately identified as “at-risk” by DIBELS was greater than the percentage of non-ELL students correctly identified as “at risk.” The opposite was true when considering students identified as “low risk.” The DIBELS assessments more accurately identified non-ELL students in the “low risk” category than ELL students in the “low risk” category. The authors concluded that, in all, the measures of oral reading fluency such as DIBELS are as effective with ELL students as with non-ELL students.

Other studies produced surprising results as well. Baker and Good (1995) found that oral reading fluency was as sensitive to growth in reading for ELL students as for English-only students. Kim (2012) found that oral language and oral reading fluency were both directly related to reading comprehension; silent reading fluency and Spanish literacy skills were not related to reading comprehension. Wiley and Good (2005) found

that oral reading fluency and the maze (timed cloze reading activity) scores were predictive of success on standardized reading assessments. Contrary to the authors' original assumptions, though, the oral reading fluency score was more predictive for ELL students than the maze comprehension assessment. The maze activity was more predictive for non-ELL students.

Chapter 3

The purpose of this chapter is to introduce the methodology used in the quantitative study comparing English Language Learners (ELL) with non-ELLs success on DIBELS and MCA benchmarks. This research will guide districts in understanding the uniqueness of ELL students and how learning to read may occur differently with these learners. Furthermore, if ELLs learn to read differently, they must also be assessed differently in order to ensure their true abilities are shown.

Research Questions and Null Hypotheses

Bottom-up reading theory starts with the smallest units of sounds and meanings and works up to larger components of comprehension and meaning (Gregory, 2016). Inversely, second language acquisition research touts the benefits of top-down reading theories which focus first on meaning and comprehension and consider smaller units of sounds and meaning as secondary (Gregory, 2016). In fact, vocabulary fluency and language development impact reading success more than a focus on phonics (Change, et al., 2020). Yet, when assessing reading progress, assessments focusing on bottom-up skills are utilized with English language learners, bringing into question their validity. Oral reading fluency (ORF) has been shown to be predictive extensively of future reading success in native English speakers (Burns, et al., 2002; Kuhn, et al., 2010; Pikulski & Chard, 2005; Yovanoff, et al., 2005).

When considering English language learners, the research is less certain. In their 2012 study, Quirk and Beem found that ELLs comprehension skills developed at a slightly lower rate than their fluency skills. Kim (2012) found that language development

played a more important part in fluency rates than did fluency instruction. Even though research shows that ELLs utilize top-down reading strategies to a greater degree and their language development has a greater influence on fluency rates, districts still employ bottom-up reading assessments such as Curriculum-Based Measures in reading (CMB-r), more specifically DIBELS, to make important decisions about reading progress and comprehension (Manzo, 2005).

DIBELS claims that their one-minute timings and corresponding benchmark cutoff scores can predict future reading success or diagnose reading difficulties (Manzo, 2005). Students read grade-level-appropriate passages for one minute after which proctors count how many words were read correctly to determine if students are on the right track to reading success (Ostayan, 2016). Often, their research excludes English language learners (Smolkowski & Cummins, n.d.) as they recognize that ELLs may not produce results equal to their native-speaking peers.

To address this discrepancy, this research seeks to find out:

1. Can DIBELS oral reading fluency (ORF) scores predict success on the Minnesota Comprehensive Assessment to the same degree as native-English speaking students in grades 3 and 4?
2. Does proficiency in the DIBELS ORF and MCA lead to proficiency on the WIDA ACCESS test?

In this study, the null hypothesis is that DIBELS ORF can equally predict reading proficiency in English learners as it does with native-English speaking

students. Additionally, ELLs who benchmark on the DIBELS ORF assessment and the MCA may not demonstrate proficiency on the WIDA ACCESS.

Research Design

This study is quantitative research. As stated in Cozby (2017), quantitative research is an effective method of evaluating programs and procedures by looking closely at the results of such programs and procedures. The author goes on to talk about the draw to using intuition and seeing correlations when, in fact, this correlation may actually be illusory. Educators, including the researcher in this study, often rely on intuition to draw conclusions and, in turn, make correlations between factors. When working with students, especially English Language Learners, it seems as though the link between reading fluency and comprehension does not exist to the extent it occurs in native-English speakers. Yet, Cozby (2017) warns that intuition is not enough. Empirical evidence must be sought to support those intuitions. Data collection and analysis can support or disprove hypotheses and support or disprove correlations.

Participants

Participants were drawn from Mankato Area Public Schools, Minnesota District #77. Because of the COVID-19 pandemic, current testing practices were put on hold for the 2019-2020 school year. States received a reprieve from standardized testing as most districts were transitioning or transitioned to distance learning during the typical testing window and test integrity would be brought into question if delivered remotely. Because of this, data was drawn from the 2018-2019 school year.

Data from students in grades three and four during the 2018-2019 school year was analyzed to determine the correlation between DIBELS ORF scores and MCA scores. Initially, the researcher looked at only collecting data from third grade but decided to include fourth grade as well. Research has shown that students tend to score higher in third grade than they do in fourth grade (McNamara, 2011). Reading in grades 3 - 5 transitions from learning to read to reading to learn. Students encounter a larger amount of expository writing and must draw on their personal experiences to comprehend what they are reading (McNamara, 2011). Many fourth graders are reading more complex, language-rich expository texts for which they have not yet had personal experience, causing them to score lower in fourth grade on formative and summative assessments. This has become known as “the fourth-grade slump.” Because of this phenomenon, the researcher added fourth-grade data to this research to see if this slump can be seen and/or predicted by DIBELS. The data set included 693 third graders and 698 fourth graders.

Instruments

Using archival data for the 2018-2019 school year, different tests were run based on DIBELS, MCA, and WIDA scores for non-ELL (all) students and ELL students. Initially, descriptive statistics were produced and analyzed for native English speakers and English learners. This data included the minimum, maximum, mean, and standard deviation for native English speakers and language learners on the MCA Reading test. Next, correlational relationships were determined using Pearson Product Moment Correlation (Pearson Correlation Coefficient or Pearson’s r). This was used to

determine the strength of the linear relationship between our two variables: the MCA Reading and the DIBELS suite of assessments. This test attempts to draw a line of best fit through the data of our two variables. The Pearson correlation coefficient shows how far away all the data points are to this line of best fit. The data was entered into SPSS to determine this line of best fit. Additionally, this same test was used to determine the relationship between MCA Reading and the ACCESS suites of assessments for English learners.

Finally, regressions were used to decide if a predictive relationship existed between the DIBELS suite of assessments and the MCA reading assessment. Regressions were chosen as they are the logarithm of odds and because our dependent and independent variables were binary or dichotomous. Simple linear regressions were used primarily to demonstrate the predictive relationship between the MCA reading assessment and each of the DIBELS assessments included in this study with the MCA reading assessment being the dependent variable and the DIBELS assessments acting as the independent variable.

Stepwise regressions were also used to show if using more than one DIBELS assessment (independent variables) can increase the predictive ability of DIBELS. This allowed us to regress multiple variables while also removing variables that were not important. This regression also allowed us to demonstrate how particular subtests, when added to the model, increased the predictive power of the tests.

Procedure

To begin, permission to use student data must be acquired. Mankato Area Public Schools requires researchers to complete the Mankato Area Public Schools Request for Approval to Conduct Research form found on their website. After consulting with the Director of Curriculum and Instruction, the researcher will complete and edit this document to the satisfaction of the Direction of Curriculum and Instruction. Once signed, the school board for Mankato Area Public Schools will need to approve the research. Once their approval is garnered, data will be requested from the District Information Systems Manager. Additionally, an Institutional Review Board (IRB) application will be completed through Minnesota State University, Mankato to ensure all aspects of the research are performed ethically and no subjects are harmed in the research. Once IRB approval has been given, data will be released to the researcher and tests will be run according to Table 1. After running all tests, the researcher will analyze the results and draw conclusions based on the results.

Chapter 4

In this chapter, the researcher will report the results of various statistical analyses to determine if a correlative relationship exists between assessments given in a local Minnesota School district. The research seeks to answer the following questions: (a) Can DIBELS ORF scores accurately predict proficiency on the Minnesota Comprehensive Assessment (MCA) for English Language Learners in grades 3 and 4 to the same degree as native-English speaking students? and, (b) Does proficiency on the DIBELS ORF and/or MCA lead to proficiency on the WIDA ACCESS test? This study provides significant implications as English learners can potentially be misdiagnosed as “poor” readers when their oral reading fluency scores are the sole measure used in determining need. It can also disguise comprehension deficiencies when the student is reading fluently enough to achieve a benchmark, giving a false impression that there are no reading needs. Educators must be able to rely on the data they get from these assessments to the same degree with English learners as they do with native English speakers.

Sample and Metrics

For this study, data was collected on all third and fourth-grade students from a rural school district in Minnesota. Because of the Coronavirus Pandemic and the inconsistency of testing within that time period, data were requested from the 2018-2019 school year. Within that sample set, students were identified as native-English speakers or English Learners (ELs). In total, data were collected for 1,392 students, 1,274 native-English speakers, and 118 English Learners. Of the 694 students in grade 3, there were

633 native-English speakers and 61 English Learners. Of the 698 students in grade 4, there were 641 native-English speakers and 57 English Learners.

Data was collected in reference to three assessments: Dynamic Indicators of Basic Early Literacy Skills (DIBELS), the Minnesota Comprehensive Assessment (MCA), and the WIDA ACCESS. Within the DIBELS suite of assessments, scores from five distinct subtests were gathered. This research focuses mainly on the DIBELS Oral Reading Fluency (ORF) Words Correct subtest and its ability to predict success on the MCA. Additionally, though, other subtests were analyzed to measure their ability to predict success as well and/or to determine if any of these subtests had greater predictive validity than the ORF Words Correct subtest. These tests included: the DIBELS Composite Spring Score, the ORF Retell score, the ORF accuracy score, and the ORF retell quality of response. The Composite Spring Score is the total score considering all of the DIBELS subtests for that grade level. In the ORF Retell, students are given one minute to retell as many words as they can from the ORF reading passage. The words are counted and that becomes their retell score. This retell is also evaluated as to its quality. For example, if the events were retold in sequential order, the retell would score higher in quality. The retell quality goes from a low score of 1 to a high score of 4. The ORF Accuracy score is calculated based on how many words the students read correctly in the ORF reading passage divided by the total number of words they attempted to read in that minute.

The MCA reading test is the standardized test that all students in grades 3-8 and 11 in Minnesota are required to take. This assessment measures student achievement

towards academic reading standards. The state of Minnesota has determined that a score of 350 in third grade and a score of 450 in 4th grade shows minimum proficiency of the academic standards.

Additionally, English Learners must also show proficiency in English on the WIDA ACCESS test. This assessment measures proficiency in listening to, speaking, reading, and writing in English. The results of this assessment are presented in two ways: scale score and proficiency level. For the purposes of this study, scale scores were garnered as these scores more precisely track student growth over time, across grade levels.

Data Analysis

MCA and DIBELS

Research question #1 seeks to determine the relationship between DIBELS ORF proficiency and MCA proficiency for English Learners in comparison to native English speakers. Is the relationship between DIBELS ORF and MCA as predictive with English Learners as with native English speakers? The data included a total of 1392 students. 694 third-grade students made up 49.7% of the total number of students whose data was sampled. 698 fourth-grade students made up 50.3% of the total number of students whose data was sampled. Descriptive statistics will first be presented for native English speakers, to be followed by English learners for each grade level. Because of the large number of assessments analyzed, and for added clarity, the data will be presented in tables. Because third grade and fourth grade MCA and DIBELS benchmarks are not the same, each will be presented separately.

Table #1*Assessment Scores for Third Grade Native English Speakers*

	N	Minimum	Maximum	Mean	Std. Deviation
MCA Reading	629	301	399	355.18	20.232
DIBELS Composite Spring	619	5	679	400.51	119.058
DIBELS ORF Words Correct	630	0	244	117.21	38.200
DIBELS ORF Retell	629	0	94	45.43	20.762
DIBELS ORF Accuracy	628	42	100	96.51	6.261
DIBELS ORF Retell Quality of Response	629	1	4	3.01	.959

Table #2*Assessment Scores for Third Grade English Language Learners*

	N	Minimum	Maximum	Mean	Std. Deviation
MCA Reading	54	301	366	339.06	15.191
DIBELS Composite Spring	54	8	506	307.44	127.935
DIBELS ORF Words Correct	54	8	166	94.37	37.723
DIBELS ORF Retell	54	0	94	37.24	20.991
DIBELS ORF Accuracy	54	40	100	92.33	10.784
DIBELS ORF Retell Quality of Response	54	1	4	2.54	1.004

When looking at the statistical differences between native English speakers and English Learners, discrepancies between achievement are readily apparent. Native English speakers achieved greater maximum scores across the three most significant assessments: the MCA Reading, DIBELS Composite, and DIBELS ORF words correct. As shown in Tables #1 and #2, the maximum score of all three of these

assessments for native English speakers is greater than that of English learners. On the MCA, native English speakers earned a mean score of 355.18 (SD = 20.232) while English Learners earned a mean score of 339.06 (SD = 15.191). Likewise, native English speakers outperformed English Learners on the DIBELS Spring Composite with an average score of 400.51 (SD = 119.058) with the English Learners scoring 307.44 (SD = 127.935). On the DIBELS ORF Words Correct, native English speakers achieved an average of 117.21 (SD = 38.200) and English learners scored 94.37 (SD = 37.723). On average, native English speakers scored 16.12 points higher on the MCA than English learners, 93.07 points higher on the DIBELS Composite, and read 22.84 more words correct. Although there are differences between native English speakers and English learners in the other DIBELS subtests, the difference is not as great as with the other assessments.

Table #3

Assessment Scores for Fourth Grade Native English Speakers

	N	Minimum	Maximum	Mean	Std. Deviation
MCA Reading	634	411	490	454.64	14.561
DIBELS Composite Spring	630	0	728	460.26	118.243
DIBELS ORF Words Correct	635	0	244	133.16	39.683
DIBELS ORF Retell	633	0	94	48.67	21.741
DIBELS ORF Accuracy	634	0	100	97.55	8.061
DIBELS ORF Retell Quality of Response	631	1	4	3.13	.912

Table #4

Assessment Scores for Fourth Grade English Language Learners

	N	Minimum	Maximum	Mean	Std. Deviation
MCA Reading	54	411	464	436.98	11.619
DIBELS Composite Spring	51	9	574	338.35	109.160
DIBELS ORF Words Correct	53	0	162	95.68	34.684
DIBELS ORF Retell	52	0	94	33.94	18.459
DIBELS ORF Accuracy	52	50	100	95.38	7.201
DIBELS ORF Retell Quality of Response	50	1	4	2.60	1.030

Similarly to third grade, fourth grade native English speakers outperformed English learners in the three main assessments: MCA Reading, DIBELS Composite Spring, and DIBELS ORF Words Correct. As shown in Tables #3 and #4, the maximum scores for native English speakers on all three assessments were greater than for English learners. Fourth-grade native English speakers averaged 454.64 (SD = 14.561) on the MCA while English Learners averaged 436.98 (SD = 11.619). On the DIBELS Spring Composite, native English speakers scored an average of 460.26 (SD = 188.243) and English learners scored 338.35 (SD = 109.160). This was particularly evident in DIBELS ORF words correct. The maximum score for a native English speaker was 244 words per minute while the maximum for English learners maxed out at 162, a difference of 82 words per minute. The mean number of words read correctly for native English speakers was 133.16 (SD = 39.683) whereas the mean number of words read correctly for English learners was 95.68 (SD = 34.684), a difference of 37.48 words per minute.

A Pearson product-moment correlation coefficient was calculated to assess the relationship between MCA Reading and all the DIBELS subtests. Of all the subtests, the DIBELS Composite Spring showed a high correlation to the MCA for both native

English speakers and English learners across both grade levels, $r(614) = .735, p < .001$ (3rd grade) and $r(622) = .729, p = .001$ (4th grade) for native English speakers) and $r(50) = .760, p < .001$ (3rd grade) and $r(48) = .705, p < .001$ (4th grade) for English learners. For native English speakers, DIBELS ORF Words Correct had the next greatest positive correlation with the MCA Reading assessment, $r(624) = .710, p < .001$ (3rd grade) and $r(626) = .729, p < .001$ (4th grade) while the correlation with English learners was not as strong, $r(50) = .577, p < .001$ (3rd grade) and $r(49) = .662, p < .001$ (4th grade). DIBELS ORF Accuracy was more correlated with MCA for English learners in third grade, $r(50) = .640, p < .001$. than for native English speakers, $r(622) = .578, p < .001$.

Interestingly, the correlation between DIBELS ORF accuracy in 4th graders and the MCA Reading was much smaller than all the other assessments with both native English speakers ($r(626) = .414, p = .000$) and English learners ($r(49) = .432, p = .000$). In 3rd grade English learners, DIBELS ORF Retell had the second highest correlation value to the MCA, $r(50) = .680, p = .000$. It was not as high with fourth grade English learners, $r(48) = .572, p = .000$. When considering DIBELS ORF Retell and native English speakers, the correlation is not as large. Although 3rd grade language learners exhibited the highest correlation, 3rd grade native English speakers demonstrated the lowest correlation, $r(623) = .449, p = .000$. 4th grade native English speakers did not show much of an improvement, $r(625) = .473, p = .000$.

After looking at correlations between MCA Reading and the DIBELS subtests, a multiple regression was used to test if each of the DIBELS subtests significantly

predicted MCA Reading scores. A multiple regression uses multiple independent variables (DIBELS assessments) to predict the outcome of a dependent variable (MCA Reading). The results of the multiple regression are reported as R-squared which represents the percent of variability in the dependent variable that can be explained by the independent variables. In this multiple regression, DIBELS ORF Words Correct showed more predictive ability with Native English speakers than with English learners, especially in third grade. DIBELS ORF Words Correct explained 50.3% (adjusted R square = .503) of the variability on the MCA for native English speakers in 3rd grade whereas it only explained 32% (adjusted R square = .320) of the variability for English Learners. In fourth grade, the percentages moved closer together with 50.8% for Native English Speakers and 42.7% for English Learners.

Figure #1

English Learners - Regression Model

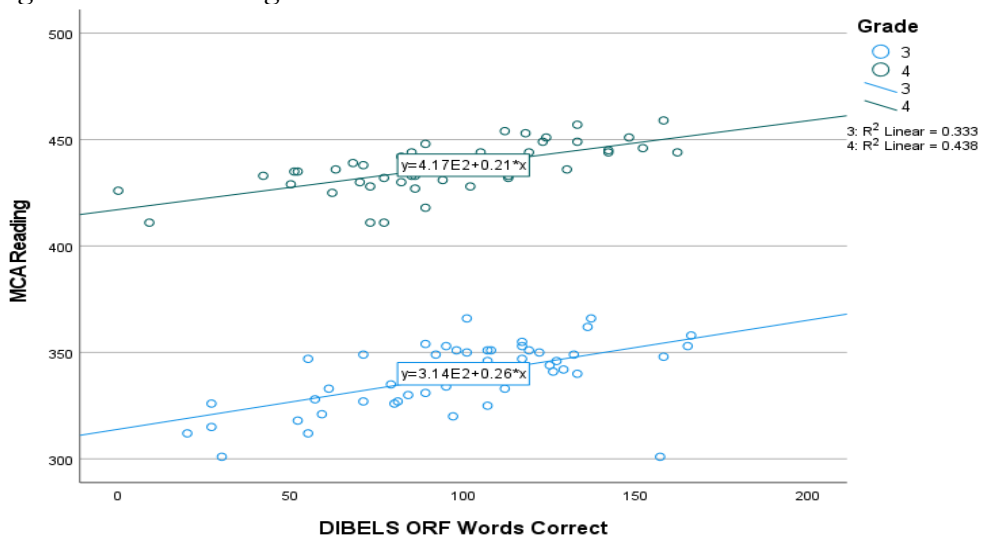
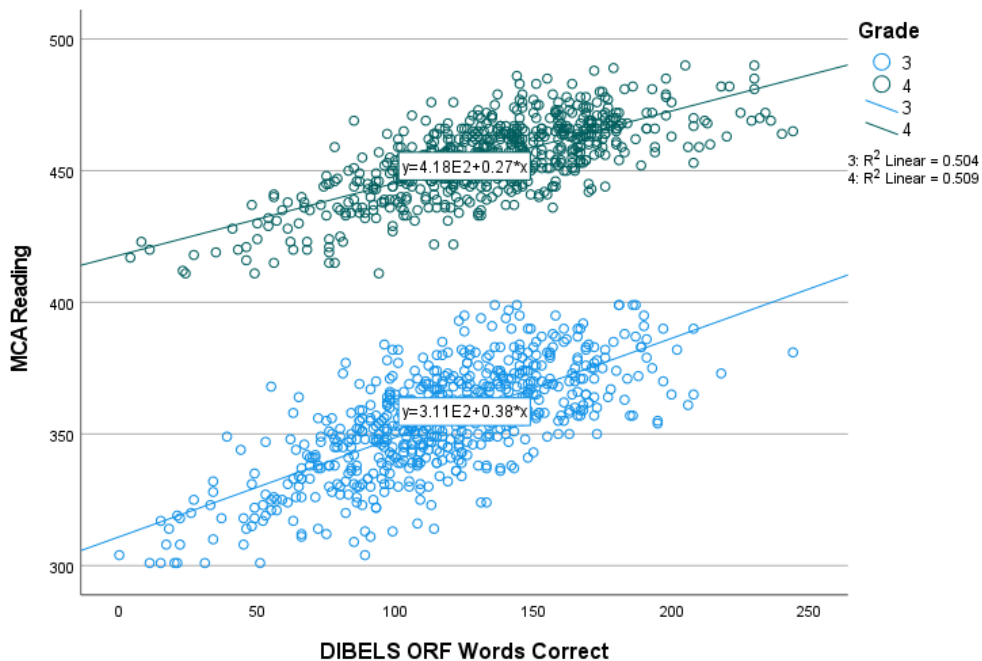


Figure #2

Native English Speakers - Regression Model



Beyond analyzing the predictive ability of the independent variables on the dependent variable, a standardized coefficient beta was used to rank the independent variables by their impact on the dependent variable. The most important independent variable will have the highest maximum absolute value. When looking at the Standardized Coefficients Beta in this research, the DIBELS assessments fared differently with Native English speakers than with English Learners. DIBELS ORF Words Correct outperformed the other DIBELS assessments for both Native English speakers and English Learners with Native English speakers outperforming English Learners. In third grade, Native English speakers show a standardized beta of .710 ($t = 25.174$, $p = .000$) and English Learners show a slightly lesser beta of .577 ($t = 4.999$, $p = .000$). As with the Adjusted R Square, the discrepancy seems to shrink as students

advance to fourth grade. Fourth grade Native English Speakers have a standardized beta of .713 ($t = 25.458$, $p = .000$) whereas English Learners have a beta of .662 ($t = 6.183$, $p = .000$). Overall, DIBELS ORF Word Correct had the highest predictive relationship of all the assessments with Native English speakers in third grade ($b = .706$, $t = 24.836$, $p = .000$) and fourth grade ($b = .710$, $t = 25.209$, $p = .000$). For English Learners, the DIBELS ORF Retell had the highest impact at third grade ($b = .680$, $t = 6.556$, $p = .000$) while the DIBELS ORF Words Correct had the highest predictive relationship in fourth grade ($b = .613$, $t = 5.314$, $p = .000$).

When considering all of the DIBELS assessments included in the study in relation to the MCA, the predictive ability of DIBELS increased with the addition of different DIBELS subtests. The independent variables (DIBELS assessments) were put into a stepwise regression that allows the independent variables to be added into the model one-by-one to determine how the additions of these variables impact the dependent variable. For Native English Speakers, the adjusted r square when looking only at DIBELS ORF Words Correct was .497. When adding DIBELS ORF Accuracy into the regression, the predictive ability went up to .521, and then to .534 when also including DIBELS ORF Retell Quality of Response. For English Learners, the adjusted r square when only looking at the DIBELS ORF Retell is .451. When considered in conjunction with the DIBELS ORF Accuracy, the adjusted r square value went up to .566.

WIDA and the MCA

The second research question asks if proficiency in the DIBELS ORF and/or MCA leads to proficiency on the WIDA ACCESS. A Pearson Correlation was used to

show the relationship between the MCA and each of the domains of the ACCESS assessment as well as the overall composite score of the ACCESS. In third grade, the ACCESS Reading showed a large correlation with the MCA ($r(53) = .695$) with the ACCESS composite being the next correlative ($r(53) = .530$). In fourth grade, the ACCESS composite showed the greatest correlation with the MCA ($r(52) = .710$) with the ACCESS composite correlated to almost the same degree ($r(52) = .704$). Interestingly, the ACCESS writing showed the lowest correlation in 3rd graders ($r(53) = .218$) whereas in fourth grade the correlation is much higher ($r(52) = .501$).

When considering the predictive power of proficiency on ACCESS with the MCA Reading, the ACCESS Reading subtest was the most predictive in both third grade ($b = .695$, $t(53) = 6.909$, $p = .000$) and fourth grade ($b = .704$, $t(53) = 7.010$, $p = .000$). But, when running a multiple regression with the ACCESS Composite and the DIBELS composite in relation to the MCA, the DIBELS composite was far more predictive than the ACCESS composite in both 3rd grade and 4th grade.

Table #5

Third Grade English Learners Stepwise Regression Model

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	264.580	17.557		15.070	.000
ACCESS Composite	.144	.058	.237	2.494	.016
DIBELS Composite Spring	.085	.012	.673	7.075	.000

a. Grade = 3

b. Dependent Variable: MCA Reading

Table #6

Fourth Grade English Learners Stepwise Regression Model

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	362.281	23.886		15.167	.000
ACCESS Composite	.165	.076	.296	2.167	.036
DIBELS Composite Spring	.050	.014	.505	3.700	.001

a. Grade = 4

b. Dependent Variable: MCA Reading

Summary

Overall, Native English speakers outperform English Language Learners across all assessments in this study. For Native English speakers in both third and fourth grade, DIBELS ORF Words Correct is correlative and predictive of the success of the MCA Reading assessment. It is less predictive for English Language Learners where the DIBELS ORF retell presented greater predictive ability. When more than one DIBELS assessment was considered, the correlation significance and the predictive ability increased for both Native English speakers and English Language Learners.

For English Language Learners, the WIDA ACCESS Reading subtest is the most predictive of success on the MCA Reading test in both third and fourth grade. Yet, the DIBELS Composite was significantly more predictive than any of the ACCESS subtests, including reading.

Limitations

Perhaps the biggest limitation of this study is the focus on only two grade levels. The MCA Reading assessment begins in third grade for Minnesota students and continues every year until eighth grade. Students then take it again in tenth grade. This research focuses only on third and fourth-grade students. Although the DIBELS Spring

Composite had the greatest correlation with both third and fourth-grade English Learners, there were differences in the correlations between third and fourth-grade English Learners and the ACCESS. There was a greater correlation between the MCA and the ACCESS Composite for third-grade English Learners than for fourth-grade English Learners. The ACCESS Writing had a higher correlation to the MCA Reading in fourth grade than it did in third grade. As educators make decisions regarding the academic abilities of and possible interventions for English Learners, it seems as though focusing on different metrics may be necessary at different grade levels. It is uncertain if these results can be generalized to include higher grade levels and/or if the results would change if other grade levels were studied.

Additionally, DIBELS assessments are given beginning in kindergarten and much research has been done to demonstrate their predictive ability for future reading success with native English speakers, less so with English learners. This research does not track students from kindergarten to third grade to make such determinations.

Chapter 5

Introduction

This chapter has four main sections. First, a summary of the research and study will be provided followed by a brief description of the findings and its conclusions. Additionally, the researcher will provide implications of the findings and recommendations for future researchers and/or educators.

Summary of the Research

Modern day educational reform has forced an era of accountability testing with additional focus placed on gaps in achievement between White students and Black, Hispanic, Asian or Pacific Islander, American Indian, English Language Learners, as well as students with disabilities and those receiving special education services (Hayes & Williams, 2008). This research focuses on the difficulty of accurately assessing the academic achievement of English Language Learners in the language for which these students have not yet reached proficiency (AERA et al., 1999, p. 91). In Minnesota, students (including English Language Learners) take the Minnesota Comprehensive Assessments (MCAs) to show academic achievement towards their grade-level standards (Minnesota Department of Ed., n.d.). Additionally, English Language Learners are required to demonstrate proficiency in English on language proficiency assessments. Minnesota has adopted the WIDA suite of assessments to evaluate language proficiency (WIDA, 2020). Language learners must show proficiency in listening, reading, speaking, and writing in English.

Because these assessments occur once per academic year, districts must use other assessments to determine progress toward grade-level standards. Oral reading fluency has become widely accepted as an effective means of measuring progress in reading comprehension and predicting success on high-stakes standardized tests (Burns et al., 2002; Yovanoff et al., 2005). The link between reading fluency and reading comprehension and overall reading achievement with English Language Learners is not as certain as with native speakers although research does show that as oral language proficiency grows, oral reading fluency also increases (Crosson & Lesaux, 2010; Kim, 2012; Quirk & Beem, 2012).

Regardless, oral reading fluency measures have become the standard for measuring growth towards academic achievement. Curriculum-based measures consisting of one-minute timings of reading passages taken from the classroom curriculum were meant to measure the effectiveness of classroom instruction and specific reading interventions (Schilling et al., 2007). DIBELS, a commercially-produced CBM, is widely used to measure growth in reading and, in turn, used to make educational decisions (Kuhn et al., 2010; Manzo, 2005). Research has shown a definitive link between DIBELS' suite of reading assessments and future reading success (Good et al., 2009). The research is not as certain when considering the unique needs of English Language Learners (Muyskens et al., 2009; Quirk & Beem, 2012). This is especially concerning knowing that many if not most educational decisions made in schools are based on oral reading fluency data.

The present study asked the questions: 1) Can DIBELS ORF scores accurately predict proficiency on the MCA for English Language Learners in grades 3 and 4 to the same degree as it can with native English speakers? and 2) Does proficiency on the DIBELS ORF and MCA lead to proficiency on the WIDA ACCESS test? Because important educational decisions are made based on this data, educators must be able to rely on the validity of the results for all learners.

Summary of the Findings

Correlational tests, multiple linear regressions, and stepwise regressions were used to determine the relationship between DIBELS and MCA for native English speakers and English language learners. When looking at the descriptive statistics, native English speakers showed means consistently higher than the means of English Language Learners on the MCA and DIBELS assessments across both grade levels.

Next correlational tests were conducted to determine if there was a general correlation between the MCA and DIBELS for Native English speakers and English Language Learners. The DIBELS composite spring score showed a strong correlation to the MCA for both native English speakers and English Language Learners. When considering the DIBELS ORF in correlation with the MCA, a stronger correlation was found for Native English speakers than with English Language Learners as has been shown in past studies. In third grade, DIBELS ORF accuracy for English Language Learners was found to have a stronger correlation with the MCA than for Native English speakers but this was not true in fourth grade. In fact, DIBELS ORF accuracy was much less correlated in 4th grade than all the other assessments with both Native English

speakers and English Language Learners. The DIBELS ORF Retell had a higher correlation for English language learners than for Native English speakers, especially in third grade.

Research question one asks if DIBELS ORF Words Correct can predict achievement on the MCA for English Language Learners to the same degree as with Native English speakers. Multiple linear regressions were used to determine the predictive validity of the DIBELS suite of assessments. For both English learners and native English speakers, the DIBELS Spring Composite had the highest correlation to the MCA Reading. The DIBELS ORF Words Correct showed more predictive ability with Native English speakers than with English Language Learners, especially third graders. In fourth grade, the discrepancy was not as statistically different. When considering all the DIBELS assessments, DIBELS ORF Words Correct was still the most predictive for Native English speakers whereas the DIBELS ORF Retell was more predictive for English Language Learners.

Research question two asks if proficiency in the ACCESS can be predicted by proficiency on DIBELS and/or the MCA. The ACCESS reading subtest had a significantly greater correlation with the MCA than the other subtests. The ACCESS Composite was almost equally correlated with the MCA as was the ACCESS reading subtest in fourth grade. This was not true in third grade, where the ACCESS Composite correlation was lower. Interestingly, the ACCESS writing subtest showed a small correlation in third grade but in fourth grade, that correlation was more significant.

Implications

Overall, this research shows that language proficiency is necessary to demonstrate proficiency on the MCAs. As students develop and grow in their language skills, their academic achievement scores increase. English Learners are often misdiagnosed with reading difficulties when given the same assessments as Native English Speakers. English learners have been identified as limited in English and are participating in English development programming (Ardasheva et al., 2012). Being identified as limited in English does not disqualify them from participating in standardized, high-stakes testing. Native English speakers outperform English Learners consistently creating the perception that English Learners are less capable than their white, non-Hispanic peers (Aradasheva et al., 2012).

This research verifies that Native English speakers outperform English Learners on the DIBELS and MCA at both grade levels. It is not surprising considering English Learners are required to test in a language for which they do not yet have proficiency. As stated in AERA et al. (1999), any test that assesses students in a language that is not their native language becomes a language assessment as well as an assessment of the competencies for which it is meant to measure. Research shows construct validity is challenged by delivering assessments in languages that are not native to the test takers and yet they are required to participate (Mahoney, 2008) and these questionable results are used punitively against districts, schools, teachers, and ultimately students. Modifications and accommodations are offered in an attempt to increase the validity but it falls short for these students. English Language Learners have an

automatic disadvantage and the results often provide a false narrative of their academic and cognitive abilities.

This is especially true of oral reading fluency measures such as DIBELS. Instructional decisions are often made solely from these one-minute timings. Students are diagnosed with reading difficulties, labeled as “at risk,” and targeted for reading interventions that focus exclusively on phonics instruction without definitively determining if phonics skills is the area of most need. This misdiagnosis and subsequent intervention can lead to phonetic reading development that advances beyond the learner’s language development. These students potentially become “word callers,” readers who decode the words but are unable to attach the appropriate level of understanding to the words (Quirk & Beem, 2012).

Additionally, Quirk and Beem (2012) discovered that “word callers” were much more likely at the intermediate level. These students may fool decision-makers into believing there are no reading difficulties as they are able to read at an appropriate fluency level yet their comprehension may be falling behind. These intermediate level students also have oral language skills approaching native-like fluency which also contributes to the belief that they do not require any sort of intervention. Valencia et al. (2010) refer to this as “false negatives.” Because of their perceived fluency, they are often categorized as “low risk” or “some risk” and disqualified from any sort of intervention. Comprehension interventions at this level may be helpful and necessary but the need may not be recognized if oral reading fluency and oral language fluency are the only measures considered.

Recommendations

This research shows that when DIBELS ORF Words Correct is used in isolation with English Language Learners, the validity is questionable. When used in conjunction with other elements of the DIBELS suite such as DIBELS ORF Retell, the validity of the results did increase. For example, if a student reached benchmark on the DIBELS ORF Words Correct and also reached the benchmark on the DIBELS ORF Retell, the results were more reliable than when considering only the DIBELS ORF Words Correct.

Relying only on the DIBELS suite of assessments still may not be enough for English Language Learners. Although the validity goes up as DIBELS assessments are added, research has shown that the sensitivity of the DIBELS suite is not as high for English Language Learners as with Native English speakers. Johnson et al. (2009) showed that an acceptable level of sensitivity for any assessment is between 90% and 95%. Valencia et al. (2010) found the sensitivity for DIBELS with English Language Learners was significantly lower than this threshold (>80%). They recommend using multiple assessments with populations who may struggle with comprehension. If DIBELS is the only option for progress monitoring English Learners, it is essential for educators and/or decision-makers to look beyond just the ORF Words Correct and evaluate the learners across all of the DIBELS assessments available.

Surprising to this research is the inability of DIBELS Accuracy to predict success on the MCA. It seemed reasonable to think that as students learned phonetic principles and applied them accurately, they would then have a greater chance of understanding the words read. Yet, DIBELS Accuracy was the least effective assessment in predicting

success for English Language Learners. This seems to indicate that English Language Learners are garnering meaning using top-down reading versus the more phonics-based bottom-up reading as indicated by the research (Krashen, 1981; Krashen & Bland, 2014; Ng et al., 2019). Educators and decision-makers would benefit from teaching and applying top-down reading strategies as well as bottom-up reading interventions with English Language Learners to enhance both abilities. Top-down strategy development will build overall language proficiency while bottom-up strategies will ensure students develop accuracy simultaneously. For example, at the classroom level, repeated readings, a top-down strategy, would allow English Learners to read passages multiple times gleaning additional meaning and vocabulary acquisition with each pass (Taguchi et al., 2012). In conjunction, English Learners could also then receive direct instruction in the “phonic approach” learning how sounds and letters combine to make words (Gregory, 2016). This balance of approaches may lead to faster language acquisition in these unique learners.

This research also indicates that language proficiency is necessary for achieving reading proficiency. The ACCESS reading subtest was a solid indicator of proficiency on the MCA reading assessment. Students who showed proficiency in the ACCESS reading were more likely to also demonstrate proficiency in the MCA reading assessment. This raises the question of whether or not both assessments are necessary. Currently, English Language Learners must make their way through the MCA reading assessment and the ACCESS suite of assessments. The Minnesota Department of Education has set limits as to how much instructional time can be used for assessments

(Minnesota Department of Education, n.d.). Because English Language Learners must complete an additional suite of assessments, they meet or exceed these guidelines. If we know that language proficiency is required for reading proficiency and we know that proficiency on the ACCESS Reading is linked to proficiency on the MCA, perhaps English Language Learners should only be required to take the ACCESS suite of assessments until they prove language proficiency. Once they have demonstrated language proficiency, the validity of the MCA reading results could be validated.

Using that score before language proficiency is attained, seems inequitable for multiple reasons, starting at the macro level and working down towards the micro-level. In Minnesota, the WIDA ACCESS requires students to demonstrate proficiency in listening, speaking, reading, and writing in English (Minnesota Department of Education, n.d.). English Language Learners are the only students in Minnesota required to demonstrate proficiency in writing. Mainstream Native English speakers do not have to prove proficiency in writing in any grade level from kindergarten through graduation. This inequity has been eased somewhat as the proficiency criteria have been loosened in recent years. Traditionally, Language Learners had to reach proficiency in all four language domains (listening, speaking, reading, and writing) before they could be considered for exiting. Oftentimes, writing proficiency was the domain that proved most difficult in which to reach proficiency. Now, English learners must reach proficiency in at least three domains and use other evidence of proficiency in the domain for which they did not demonstrate proficiency on the ACCESS test. Other evidence could include classroom artifacts (Minnesota Department of Education, n.d.).

From a systems level in Minnesota, the academic efficacy of districts and schools is judged on MCA reading and math scores. Because of this, these subjects being tested have become the main focus in educational systems, essentially eliminating subjects outside of reading and math (Neill, 2003). Some of these subjects could support and build up the language acquisition of English Language Learners. Language associated with science, social studies, and other content areas has been significantly reduced at the elementary levels in favor of intensifying math and reading instruction, essentially teaching only to the test (Smyth, 2008). Language learners need exposure to a wide variety of academic disciplines and languages in order to develop the proficiency needed for academic success in English. Accountability testing has essentially eliminated the very language support English learners need in order to perform on those same accountability measures.

Ironically, the very subjects that have been significantly reduced or eliminated in elementary classrooms are the very subjects used in assessing English language proficiency. The WIDA ACCESS assesses the language associated with Language Arts, Mathematics, Science, and Social Studies (WIDA, 2007). English Learners are exposed mainly to the language of Language Arts and Mathematics in the mainstream classroom yet must also show proficiency in the language associated with Science and Social Studies. Native English speakers do not have to demonstrate this same proficiency in these content areas. The MCA Reading and Math tests are the only proficiencies they are required to demonstrate. Perhaps, if language learners were allowed to focus on language acquisition prior to being held accountable for academic achievement, systems could

support the diverseness of language acquisition, intensify linguistic support, and/or accelerate acquisition.

Perceptions about the ability of language learners are also an inequity that stems from assessing them in English before they are proficient. Educators are led to believe that language learners might be lagging behind cognitively. Bias is developed within educators about the capabilities of English learners and often expectations and rigor are lowered in response. Data is rarely shared about the academic achievement of Language learners who have reached proficiency and have exited language development programs because they are not reported as language learners anymore (Aradasheva et al., 2012). Educators only hear about those still making their way through the language instruction programs and do not get to follow those who exit and are successful. By waiting until language proficiency has been attained before assessing for academic achievement in English, educators could address their biases surrounding the cognitive abilities of their learners and focus attention on developing their language skills.

Perhaps the most concerning inequity for our English Learners is their belief in their own ability. No matter how little English they have acquired, they must sit in front of an academic assessment written in a language that is not written in their native language and attempt to prove themselves. In Minnesota, Recently Arrived English Learners (RAELs) must also attempt the MCA, many of whom have only been learning English for a very short time (Minnesota Department of Education, n.d.). From the start, our educational system proves to them that they do not measure up to the standards we have set. Again, if we move to focus on assessing language proficiency only until they

reach proficiency levels, we no longer dwell on what the students are not able to do but instead shift to what they truly need to achieve academic success. In turn, students may not immediately see themselves as failing but begin to see growth towards the language proficiency that will lead them to academic success.

Future research could look at many of the inequalities pointed out to determine if native English speakers could demonstrate proficiency in all the areas required of English learners, more specifically in the areas of writing and content-specific proficiency. Although writing is taught at the elementary level, the focus on writing has also diminished since the advent of standardized testing in reading and math. Could native English speakers reach the same proficiency level required of language learners? Additionally, if explicit science and social studies instruction has become less common, can native English speakers demonstrate proficiency in these areas to the same degree required of English learners? If requiring more of our English learners is the norm, we need to ensure the proficiency expectations are attainable and equal to the expectations we have of Native English speakers.

References

- Aldhanhani, Z. & Abu-Ayyash, E. (2020). Theories and research on oral reading fluency: What is needed? *Theory and Practice in Language Studies*, 10(4), 379-388.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
- Aljumah, F.H. (2020). Second language acquisition: A framework and historical background on its research. *English Language Teaching*, 13(8), 200-207.
- Amrein, A., & Berliner, D. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership*, 60, 32-38.
- Anderson, R. (1994). *Role of the reader's schema in comprehension, learning, and memory*. Ruddell, Ruddell, and Singer.
- Ardasheva, Y., Tretter, T.R., & Kinny, M. (2012). English language learners and academic achievement: Revisiting the threshold hypothesis. *Language Learning*, 62(3), 769-812.
- Ardoin, S., Christ, T., Morena, L., Cormier, D., & Klingbeil, D. (2013). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology*, 51, 1-18.

- Ausubel, D. (1968). *Educational psychology: A cognitive view*. Holt, Rinehart, and Winston.
- Bailey, A.L., & Carroll, P.E. (2015). Assessment of English language learners in the era of new academic content standards. *Review of Research in Education*, 39, 253-294.
- Baker, S., & Good, R. (1995). Curriculum-based measurement of English reading with bilingual Hispanic students: a validation study with second-grade students. *School Psychology Review*, 24(4).
- Baker, D., Stoolmiller, M., Good, R., III, & Baker, S. (2011). Effect of reading comprehension on passage fluency in Spanish and English for second-grade English learners. *School Psychology Review*, 40(3), 331-351.
- Betts, J., Reschly, A., Pickart, M., Heistad, D., Sheran, C., & Marston, D. (2008). An examination of predictive bias for second grade reading outcomes from measures of early literacy skills in kindergarten with respect to English-language learners and ethnic subgroups. *School Psychology Quarterly*, 23(4), 553-570.
- Buck, J., & Torgesen, J. (2003). The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test (FCRR Technical Report No. 1). Tallahassee, FL: Florida Center for Reading Research.
- Bunch, M.B. (2011). Testing English language learners under No Child Left Behind. *Language Testing*, 28(3), 323-341.

- Burns, M.K., Tucker, J.A., Hauser, A., Thelen, R.L., Holmes, K.J., & White, K. (2002). Minimum reading fluency rate necessary for comprehension: A potential criterion for curriculum-based assessments. *Assessment for Effective Intervention*, 28(1), 1-7.
- Chang, Y., Taylor, J.S.H., Rastle, K., & Monaghan, P. (2020). The relationship between oral language and reading instruction: Evidence from a computational model of reading. *Cognitive Psychology*, 123.
- Chi, Y., & Lin, C. (2012). Alignment between the Minnesota Academic Standards in Mathematics and Science, and the WIDA Consortium's English Language Proficiency Standards: a report on research carried out by the University of Illinois on behalf of the WIDA Consortium. Minnesota Standards Alignment Report.
- <https://wida.wisc.edu/sites/default/files/Website/State%20Pages/Minnesota/MN%20Correspondence%20Report%20final.pdf>
- Cozby, P. (2017). *Methods in behavioral research*. McGraw-Hill.
- Crosson, A.C., & Lesaux, N.K. (2010). Revisiting assumptions about the relationship of fluent reading to comprehension: Spanish-speakers' text-reading fluency in English. *Reading and Writing*, 23(5), 475-494.
- Cutting, L.E., & Scarborough, H.S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10(3), 277-299.

- Deatz, R., Smith, E., Thacker, A., Dickinson, E., Levinson, H., & Nemeth, Y. (2013). Independent review of the Reading Minnesota Comprehensive Assessment (MCA-III). Human Resources Research Organization (HumPRO), Virginia.
- Fewster, S., & Macmillan, P. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial and Special Education, 23*(3), 149-156.
- Friesen, D.C. & Haigh, C.A. (2018). How and why strategy instruction can improve second language reading comprehension: A review. *Reading Matrix: An International Online Journal, 18*(1), 1-18.
- Fuchs, L.S., & Deno, L.S. (1991). Paradigmatic distinction between instructionally relevant measurement models. *Exceptional Children, 57*, 488-500.
- Fuchs, L.S., Fuchs, D., Hosp, M.K., & Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 239-256.
- Goffreda, C.T., Diperna, J.C., & Pedersen, J.A. (2009). Preventative screening for early readers: Predictive validity of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). *Psychology in the Schools, 46*(6).
- Goldenberg, C. (2008). Teaching English language learners: What the research does and does not say. *American Educator, 33*(2), 8-23.

- Good, R.H., Gruba, J., & Kame'enui, E.J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high stakes outcomes. *Scientific Studies of Reading*, 5(3), 257-288.
- Good III, R., Kaminski, R.A., Smith, S., Simmons, D., Kame'enui, E., & Wallin, J. (2002). Reviewing outcomes: Using DIBELS to evaluate a school's core curriculum and system of additional intervention in kindergarten. *Reading in the classroom: Systems for observing teaching and learning*. Baltimore: Paul H. Brookes.
- Good, R.H., Simmons, D., & Kame'enui, E. (2009). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, 5(3), 257-288.
- Gregory, E. (2016). Learning to read: A third perspective. *Prospects*, 46, 367-377.
- Hintz, J.M., & Shapiro, E.S. (1997). Curriculum-based measurement and literature-based reading: Is curriculum-based measurement meeting the needs of changing reading curricula? *Journal of School Psychology*, 35, 351-375.
- Kim, Y. (2012). The relations among L1 (Spanish) literacy skills, L2 (English) language, L2 text reading fluency, and L2 reading comprehension for Spanish-speaking ELL first grade students. *Learning and Individual Differences*, 22, 690-700.

- Kim, J.S., Vanderwood, M.L., & Lee, C.Y. (2016). Predictive validity of curriculum-based measures for English learners at varying English proficiency levels. *Educational Assessment, 21*(1), 1-18.
- Kim, Y., Wagner, R., & Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from grade 1 to grade 2. *Journal of Experimental Child Psychology, 113*, 93-111.
- Krashen, S. (1981). *Second language acquisition and second language learning*. University of Southern California.
- Krashen, S.D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.
- Krashen, S. & Bland, J. (2014). Compelling comprehensible input, academic language and school libraries. *Children's Literature in English Language Education, 2*(2).
- Kuhn, M.R., Schwanenflugel, P.J., & Meisinger, E.B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly, 45*(2), 230-251.
- LaBerge, D. & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323.
- Lazarin, M. (2006). Improving assessment and accountability for English language learners in No Child Left Behind Act. (Issue Brief No.6). Washington, DC: National Council of La Raza.

- Liu, K.K., & Anderson, M. (2008). Universal design considerations for improving student achievement on English language proficiency tests. *Assessment for Effective Intervention*, doi:10.1177/1534508407313242.
- Mahoney, K. (2008). Linguistic influences on differential item functioning for second language learners on the National Assessment of Educational Progress. *International Journal of Testing*, 8:1, 14-33.
- Manzo, L. (2005). National clout of DIBELS test draws scrutiny. *Education Week*, 25(5), 1-12.
- McNamara, D. S., Ozuru, Y., & Floyd, R. G. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education*, 4(1), 229-257.
- Metsala, J. L., & Ehri, L.C. (1998). *Word recognition in beginning literacy*. Lawrence Erlbaum Associates, Inc.
- Minnesota Department of Education (n.d.). Accountability indicators. <https://education.mn.gov/MDE/dse/account/indicators/>
- Minnesota Department of Education Student Support Division. (2017). Minnesota Standardized English Learner Procedures. <https://education.mn.gov/mdeprod/groups/educ/documents/hiddencotent/bwrl/mdcy/~edisp/mde072039.pdf>
- Muyskens, P., Betts, J., Lau, M.Y., & Marston, D. (2009). Predictive validity of curriculum-based measures in the reading assessment of students who are English language learners. *The California School Psychologist*, 14, 11-21.

- National Commission on Excellence in Education (1983). *A nation at risk: The imperatives for educational reform*. Washington, D.C.: US Department of Education.
- Ng, Q.R., Renandya, W.A., & Chong, M.Y.C. (2019). Extensive reading: Theory, research, and implementation. *TEFLIN Journal*, 30(2).
- Nunan, D. (1999). *Second language teaching and learning*. Heinle and Heinle.
- Ostayan, J. (2016). Early literacy skills and English language learners: An analysis of students in a Title I school. *Reading Psychology*, 37, 1097-1118.
- Pikulski, J.P., & Chard, D.J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, 58(6), 510-519.
- Quirk, M., & Beem, S. (2012). Examining the relations between reading fluency and reading comprehension for English language learners. *Psychology in the Schools*, 49(6), 539-553.
- Reyes, P. & Rorrer, A. (2001). US school reform policy, state accountability systems and the limited English proficient student. *J. Education Policy*, 16(2), 163-178.
- Schilling, S., Carlisle, F., Scott, S., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *The Elementary School Journal*, 107(5), 429-448.
- Shaw, R., & Shaw, D. (20021). DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP). (Technical Report). Eugene, OR: University of Oregon.

- Sheffel, D., Lefly, D., & Houser, J. (2012). The predictive utility of DIBELS reading assessment for reading comprehension among third grade English language learners and English speaking children. *Reading Improvement, 49*, 75-92.
- Sireci, S.G., Han, K.T., & Wells, C.S. (2008). Methods for evaluating the validity of test scores for English language learners. *Educational Assessment, 13*, 108-131.
- Smith, F. (1994). *Understanding reading* (5th ed.). Lawrence Erlbaum.
- Taguchi, E., Gorsuch, G., Takayasu-Mass, M., & Snipp, L. (2012). Assisted repeated reading with an advanced-level Japanese EFL reader: A longitudinal diary study. *Reading in a Foreign Language, 24*, 30-55.
- Valencia, S.W., Smith, A.T., Reece, A.M., Li, M., Wixson, K.K., Newman, H. (2010). Oral reading fluency assessment: Issues of construct, criterion, and consequential validity. *Reading Research Quarterly, 45*(3), 270-291.
- Vanderwood, M.L., Tung, C.Y., & Checca, C.J. (2014). Predictive validity and accuracy of oral reading fluency for English learners. *Journal of Psychoeducational Assessment, 32*(3), 249-258.
- Vander Meer, C.D., Lentz, F.E. & Stollar, S. (2005). The relationship between oral reading fluency and Ohio proficiency testing in reading. (Technical Report). Eugene, OR: University of Oregon.
- Villanueva de Debat, E. (2006). Applying current approaches to the teaching of reading. *English Teaching Forum*.
- Welsh, J (2003). Minnesota school testing, accountability plan gets federal approval. *Knight Ridder Tribune Business News*.

- Wiley, H., & Deno, S. (2005). Oral reading and maze measures as predictors of success for English language learners on a state standards assessment. *Remedial and Special Education, 26*, 207–214.
- Wise, J., Sevik, R., Morris, R., Lovett, M., Wolf, M., Kuhn, M., Meisinger, B., & Schwanenflugel, P. (2010). *The relationship between different measures of oral reading fluency and reading comprehension in second-grade students who evidence different oral reading fluency difficulties*. *Language, speech, and hearing services in schools, 41*(3), 340-348.
- Wolf, M.K., Farnsworth, T., & Herman, J. (2008). Validity issues in assessing English language learners' language proficiency. *Educational Assessment, 13*, 80-107.
- Yesli-Dagli, U. (2010). Predicting ELL students' beginning first grade English oral reading fluency from initial kindergarten vocabulary, letter naming, and phonological awareness skills. *Early Childhood Research Quarterly, 26*, 15-29.
- Yovanoff, P., Duesbury, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24*(3), 4-12.

