MONITORING MOTIVATION AND ACADEMIC GROWTH IN WRITING FOR YOUNG ENGLISH LANGUAGE LEARNERS

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Robert Alexander Smith

Dr. Erica Lembke, Dissertation Supervisor

MAY 2018

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

MONITORING MOTIVATION AND ACADEMIC GROWTH IN WRITING FOR YOUNG ENGLISH LANGUAGE LEARNERS

presented by Robert Alex Smith, a candidate for the degree of doctor of philosophy, and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Erica Lembke

Dr. Melissa Stormont

Dr. Kristen McMaster

Dr. Stephen Kilgus

DEDICATION

This dissertation is dedicated to my ever-supportive family: to my wife, for dragging me into this whole thing to begin with, perhaps only to accompany her as she did it as well; to my oldest son Kadri, who managed throughout so many changes in lifestyle and even helped out on occasion; and to my youngest son Zane, who decided to join us toward the end of the journey just to make a bit more fun. Finally, it is also dedicated to my grandfather, who passed away only months before this dissertation was completed. Robert H Smith Jr. was a loving husband, father, grandfather, and great grandfather. His life and passing served to remind me of the importance of family and of just trying your best to be a good and honest person that was willing to work hard without forgetting to enjoy life. Although I doubt he would have enjoyed reading this document, he certainly would have been proud that I wrote it. Below are three quotes from his favorite author, Zane Grey, that I think summarize his essence and serve as my model for being:

"I am tired. My arm aches. My head boils. My feet are cold. But I am not aware of any weakness"

"Love grows more tremendously full, swift, poignant, as the years multiply"

"Today I began the novel that I determined to be great."

ACKNOWLEDGEMENTS

I must begin by thanking my mentor and advisor, Dr. Erica Lembke, for your support and guidance throughout my time at Mizzou. You may be the first person in higher education to see a potential in me that you actually wanted to work with, for better or worse. I have learned a great deal about being a scholar and researcher from you and will always strive to meet the high standards you expect from me. I know that I may not be the best at voicing my gratitude but I am forever in your debt. Thanks also to my amazing committee members. Dr. Melissa Stormont, you taught one of my first classes and did not threaten to kick me out even once. Your optimism and kindness ushered me through that difficult first semester and nothing pleases me more than to have you as a part of the final stage as well. Dr. Stephen Kilgus, you were the first person I met at Mizzou via Skype while interviewing for a different program. Although I did not end up in the program with you, the interview left me knowing that I would be joining a college led by skilled researchers who shared similar interests. Dr. Kristen McMaster, I aspire to be a researcher with your level of precision and ability to communicate with clarity and meaning. In summary, I not only have one of the best collections of researchers mentoring me, but I also have one of the best collections of people mentoring me.

I would also like to thank a host of other faculty members that have provided support and guidance across my time at Mizzou. Dr. Cathy Thomas, for teaching me the ins and outs of a good literature review and bringing me food when I was sick. Dr. Chad Rose, for taking the time to try and teach me advanced statistics and always being willing to involve me in projects. Dr. Delinda Van Garderen, for challenging and pushing me on a theoretical level while always maintaining a supportive position. Sometimes the only

way to grow is to be questioned and no one does it better than you Dr. VG. In deed, all of my interactions with faculty at Mizzou have promoted my growth as a scholar and human being.

This dissertation would not have been possible without the help from many colleagues. Thank you Carol Garmon and Kim Moore for being so knowledgeable, helpful, and all around awesome to work with. A big shout out to Erica Mason and Mary Decker for helping when you had no absolutely no reason to and a willingness to debate with me on everything from the best pizza in town to disability studies. I know that debating me can be tiring. A big thanks to Matthew Peterson for helping to collect and clean data as well as for providing me the opportunity to mentor you. I learned a great deal from you and hope that you gained a bit from me. To Kate Sadler, we came into this together and despite all odds, we are finishing it together! To Drs. Abby Allen and Apryl Poch, for being my colleagues and mentors. Abby, I would have never gotten the bloody table of contents without your template. Naturally, none of this would have been possible without the support of Columbia Public Schools, its ESL teachers, and all of the amazing students. I would like to especially thank Shelly Fair for being so supportive and advocating for the project.

Finally, I must acknowledge my loving and supportive family. My parents, DiAnne and Dr. Bob Smith, thank you for all of the baby sitting and getting the brunt of my anger and anxiety during the most stressful times. Thanks to my sister, Ashley O'Neil, and her family for always sending popcorn and wine when it was needed most. Kadri, my oldest son, I thank you for reminding me that dinosaurs are indeed cool. Zane, my youngest son, thank you for being such a happy little dude. Finally, a big thanks to

my wife, Erin Smith. I do not know that I would have ever been capable of finishing this without you. We have been around the world, lost shoes on mountaintops, been eaten alive by mosquitoes in the jungles of Borneo, changed a poop covered baby in the middle of no-where Oman, braved driving the roads of Abu Dhabi, been on a plane together that was struck by lightening in Chicago, and been stuck in Milan with no place to sleep on a holiday when everything was closed. This PhD thing though, it's definitely at the top of the 'challenges we've faced together' list. Especially since we absurdly decided to do it at the same time and have another child during the process as well. Leave it to us to add degrees of difficulty to something many consider impossible to start with. Now, let's have a glass of good wine and talk about our next adventure.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii
LIST OF TABLES vii
GLOSSARY ix
ABSTRACT xi
CHAPTER 1: INTRODUCTION AND CONCEPTUAL FRAMEWORK 1
Problem Statement 1
Purpose 3
Research Questions 4
Conceptual Framework 4
CHAPTER 2: REVIEW OF THE LITERATURE 6
The Current State and Characteristics of Education for ELs 6
Key Variables in an EL's Literacy Development 7
Early Identification of Risk and Disability with ELs. 16
Summary and Conclusion 26
CHAPTER 3: METHODS 29
Research Questions 29
Participants and Setting 30
Predictor Variables 33
Criterion Variables 37
Procedures 40
Data Analysis 45
CHAPTER 4: RESULTS 50
Descriptive Statistics 50
Reliability 61
Criterion Validity 63
Sensitivity to Growth 72
EL Performance Compared to the General Population 87
Integration of CBM-W and SAEBRS-AB to Predict Academic Performance 90
CHAPTER 5: DISCUSSION 108
Overview 108

Technical Adequacy of CBM-W 108
Utility of Combining CBM-W and Academic Behavior 116
Limitations 120
Implications for Practice 122
Future Research 122
REFERENCES 125
APPENDIX 125
Teacher Consent Form 146
Parent Consent 149
Student Assent 151
Word Dictation 152
Picture Word 153
SAEBRS-AB 154
Word Dictation Administration Directions 155
Picture Word Administration Directions 156
Picture Word Scoring Guide 161
Fidelity of Administration for Word Dictation 167
Fidelity of Administration for Picture Word 169
VITA 171

LIST OF TABLES

Table

1. Participant School Demographics……..…….....39
2. Student Participant Demographics…….....41
3. Descriptive Statistics for ACCESS ……...58
4. ACCESS 1st Grade Descriptive Statistics with Outlier Remove……...59
5. Correlations between ACCESS Sub-Tests……60
6. Descriptive Statistics and Inter-Correlations for MAP…….……61
7. Descriptive Statistics for Fall WD……..…….…….....62
8. Descriptive Statistics for Winter WD……..……..63
9. Descriptive Statistics for Spring WD……64
10. Descriptive Statistics for Fall PW…….65
11. Descriptive Statistics for Winter PW……66
12. Descriptive Statistics for Winter PW……67
13. Means and Standard Deviations of the SAEBRS…….68
14. Rates of Risk According to SAEBRS……..68
15. Internal-Consistency of SAEBRS Sub-Scales…….69
16. Alternate Form Reliability for Word Dictation……....70
17. PW Alternate Form Reliability…….71
18. 1st Grade Concurrent Validity Correlations to ACCESS-W…….72
19. 2nd Grade Concurrent Validity Correlations to ACCESS-W……73
20. 3rd Grade Concurrent Validity Correlations to ACCESS-W…….73
21. 1st Grade Predictive Validity Correlations for Fall Predictors to ACCESS-W….75
22. 2nd Grade Predictive Validity Correlations for Fall Predictors to ACCESS-W....75

23. 3rd Grade Predictive Validity Correlations for Fall Predictors to ACCESS-W….76

24. Divergent Validity of Predictors to ACCESS-LC and ORC…….77

25. Correlations Between Predictors and MAP-ELA for 3rd Grade……80

26. Divergent Validity Between MAP-ELA and MAP-MA for 3rdGrade……...81

27. Divergent Validity Between MAP-ELA and MAP-MA for 3rd Grade.....82

28. Mean & Standard Deviation of PW-WSC & PW-CWS Across Grades/Times…83

29. Comparison of Mean EL and General Population on WD……99

30. Comparison of Mean EL and General Population on PW……...100

31. Hierarchical Regression Predicting ACCESS-W from Winter WD-CLS……...103

32. Hierarchical Regression Predicting ACCESS-W from Fall WD-CLS……104

33. Hierarchical Regression Predicting ACCESS-W From Winter PW-WSC…….106

34. Hierarchical Regression Predicting ACCESS-W from Fall PW-WSC……...108

35. Hierarchical Regression Predicting MAP-ELA from Winter WD-CLS…….109

36. Hierarchical Regression Predicting MAP-ELA from Fall WD-CLS……..111

37. Hierarchical Regression Predicting MAP-ELA from Winter PW-WSC…….112

38. Hierarchical Regression Predicting ACCESS-W from Fall PW-WSC……...113

39. Numbers of Participants “at-risk” on ACCESS-W……..…115

40. Logistic Equation Results for Fall Predictors to ACCESS-W…….115

41. ROC Curve Analysis with Predictors and ACCESS-W……..116

42. Logistic Regression of Combined Predictors to MAP-ELA……....118

43. ROC Curve Analysis with Predictors and MAP-ELA…….119

GLOSSARY

English Language Learner (EL): any school age student who speaks a language other than English in the home and their home language impacts their ability to meet proficiency on state academic assessments, participate fully in instruction delivered in English, and participate fully in society. In this study, an EL must be identified as needed English as Second or Other Language services (ESOL) by the district. Also referred to as Limited English Proficient (ELP), Dual-Language Learners (DLL), or Emergent Bi-Lingual (EBL), or among other terms, if various states and in various literature.

English Language Proficiency (ELP): ELP is the students current skills in English across the domains of receptive oral English, expressive oral English, English reading, and English writing. No Child Left Behind (NCLB, 2006) required states to assess and report ELP for ELs. ELP is assessed using a variety of tests with the Assessing Comprehension and Communication in English State-to-State English Language Proficiency (ACCESS) test being the most common across states. ELP is often reported as *Beginning, Intermediate, Advanced,* or *Monitored*.

English as Second or Other Language Supports (ESOL): ESOL constitutes specialized instruction targeting language development and academic content supports for those identified as EL. Typically, ESOL supports are delivered by a trained teacher of ESOL (TESOL) using pullout, push-in, or collaborative supports. Some ESOL supports are English-on, in which instruction is delivered only in English, while others include delivering content and instruction in the EL's native language (L1) to varying degrees.

Motivation: motivation constitutes several theories all involving cognitive processes that energize behavior.

Motivated Academic Behavior: observable behavior that has been energized via motivation within an academic setting. These behaviors include task engagement, task persistence, task completion, and receptiveness to feedback. Such behavior results from high self-efficacy, outcome expectancy, and values related to the academic skills being taught/practiced.

Technical Adequacy: is the degree to which an assessment is reliable and valid.

Reliability: is the extent to which a measure produces the same outcome across time and forms.

Validity: is the extent to which a measure actually measures the construct it intends to measure. Validity can be established by analyzing the degree to which the measure in question is related to another measure that has been shown to be a quality measure of the construct, sometimes referred to as a criterion or gold-standard measure.

Content Validity: is established by examining the questions contained in an assessment to determine the extent to which they relate to the content of the construct being measured.

Divergent Validity: is based upon the premise that a measure should predict performance in the domain being assessed but not in another unrelated domain. For example, a reading assessment should be more predictive of performance in reading than in math.

MONITORING MOTIVATION AND ACADEMIC GROWTH IN WRITING FOR YOUNG ENGLISH LANGUAGE LEARNERS

Robert Alexander Smith

Dr. Erica Lembke, Dissertation Supervisor

ABSTRACT

The purpose of this study was to explore the technical adequacy and appropriateness of using benchmarks established with the general population with two forms of Curriculum Based Measures-Writing (CBM-W), Word Dictation (WD) and Picture Word (PW), with English Language Learners (ELs) in the 1st through 3rd grades as well as explore the utility of combining a measure of motivated academic behavior (i.e., Social, Academic, and Emotional Behavior Risk Screener-Academic Behavior sub-scale (SAEBRS-AB)) with CBM-W for identifying risk in writing for young ELs. ELs in the 1st through 3rd grades (n = 71) were administered two forms of WD and PW in the fall, winter, and spring of the same academic year. Teachers (n = 9) also completed the SAEBRS-AB at each time-point for each participating student. Correlations between forms at each time-point were used to establish alternate form reliability and validity was established using two criterion measures via correlations and regression. The utility of combining CBM-W with SAEBRS-AB was examined via logistic regression and Receiver Operator Characteristic (ROC) curve analysis using researcher determined cut-scores for risk on the two criterion measures. Results indicated that both forms of CBM-W are reliable and valid measures of general writing performance for young ELs, that benchmarks drawn from the general population are generally applicable to young ELs, and that integrating the SAEBRS-AB with either form of CBM-W improves diagnostic accuracy.

CHAPTER 1

INTRODUCTION AND CONCEPTUAL FRAMEWORK

Problem Statement

Writing is critical to literacy development and influences outcomes across multiple aspects of an individual's personal, academic, and professional life (Abbott, Berninger, & Fayol, 2010; Berninger & Abbott, 2010; Graham & Herbert, 2011; Graham & Perin, 2007; Mueller & Oppenheimer, 2014). Writing is also a complex literacy skill with high language demands that is innately personal and considered a reflection of one's own social and cultural identity (Bazerman, 2016; Danzak & Silliman, 2014; Ferdman, 1990). Thus, linguistic, academic, and affective variables combine to make English writing in the U.S. education system particularly difficult for many English learners (ELs). The Department of Education defines EL as any school age student who speaks a language other than English in the home and whose home language impacts their ability to meet proficiency on state academic assessments, participate fully in instruction delivered in English, and participate fully in society (Burr, Haas, & Ferriere, 2015). The academic performance of ELs is generally behind that of their non-EL peers, but this discrepancy is even wider for tasks with higher levels of language demand, such as writing (Abedi & Gandara, 2006). However, effective assessment for risk and early intervention have been shown to improve student outcomes in writing (Berninger & Amtmann, 2003; Graham, Harris, & Larsen, 2001).

Curriculum based measures in writing (CBM-W) is one type of promising writing assessment for all students, including ELs (Keller-Margulis, Payan, Jaspers, & Brewton, 2016). CBM is a global indicator of academic skill within a specific domain that can be

used to screen for risk, monitor student progress across time, and has standardized administration and scoring rules allowing for comparison within and across classrooms as well as grades (Deno, 1985; Deno et al., 2009). However, ELs represent a heterogeneous group of students with unique characteristics and needs and the generalization of benchmarks and rates of growth drawn from non-ELs is not recommended (Abedi & Garanda, 2006; Burr et al., 2015). Few studies have directly explored the reliability and validity of CBM-W with ELs specifically. Therefore, it is important to validate measures for young ELs based upon the performance of other ELs using socially valid criterion measures.

One potential issue in identifying risk amongst young ELs using static scores generated by CBM in the fall is that ELs come to school with a wide range of prior academic experiences and exposure to both oral English and English text, which may impact the predictive validity of static scores. One of the hypothesized benefits of CBM in identifying risk for young ELs is the ability to monitor progress, or responsiveness to instruction, over time. This process of progress monitoring could limit the number of false positives for risk when an EL's initial performance is a function of a lack of prior opportunities to learn rather than an indication of the individual's lack of responsiveness to prior instruction. However, reliance upon progress monitoring to confirm risk may still result in a delay of access to support for those who are at-risk. Fortunately, research with the general population has shown that motivation is also an important predictor of future writing performance that may be influenced less by prior opportunities to learn than CBM static scores (Graham, Berninger, & Fan, 2007; Pajares, 2003). For example, an EL recently arrived from Syria may be motivated to learn to write English but score poorly

on CBM-W because they speak little English and have had limited opportunities to learn to write in English. Their motivation indicates a high potential to take advantage of future opportunities to learn where CBM-W indicates high levels of initial risk. Motivated academic behavior, a common proxy measure for motivation, is behavior that increases the number and quality of opportunities to learn (e.g., task engagement, task persistence, tack completion, receptiveness to feedback) while unmotivated academic behavior (e.g., off-task, lack of task persistence, task incompletion, non-responsiveness to feedback) decreases the number and quality of opportunities to learn (DiPerna, 2006; DiPerna, Volpe, & Elliott, 2001; Meece, Anderman, & Anderman, 2006). Motivated academic behavior sets the stage for future learning. Therefore, the integration of data related to academic behaviors associated with motivation (e.g., on-task, task engagement, task persistence) and CBM-W data may improve the predictive validity and diagnostic accuracy of CBM-W.

Purpose

The purpose of this study was to examine the technical adequacy of two promising forms of CBM-W, word dictation (WD) and picture word (PW), for ELs in the 1st-3rd grades. Specifically, the study examined the inter-scorer reliability, alternate form reliability, and criterion validity of WD and PW across various time-points (fall, winter, spring) using the criterion measures of the Assessing Comprehension and Communication in English State-to-State English Language Proficiency writing subtest (ACCESS-W), a commonly used English Language Proficiency (ELP) assessment, and the Missouri Assessment Program-English Language Arts (MAP-ELA) assessment. The study also explored evidence relating to WD and PW's sensitivity to growth over time for

ELs and compared EL performance to that of the general population on WD and PW. Finally, this study examined the impact of a measure of motivated academic behavior in writing upon the predictive capabilities and diagnostic accuracy of both PW and WD.

Research Questions

1. What is the technical adequacy of WD and PW as indicators of general writing performance for ELs in the 1st-3rd grades?
    a. What is the inter-scorer and alternate form reliability of WD and PW?
    b. What is the concurrent and predictive validity of WD and PW with the criterion measures?
    c. What is the sensitivity to growth of WD and PW for ELs?
    d. How do WD and PW performance by ELs compare to the general population?
2. How does the inclusion of a measure of motivated academic behavior in writing impact the predictive validity and diagnostic accuracy of WD and PW for ELs?

Conceptual Framework

Prevention science is a model applied in many fields and arose in the late 1990s in response to a need for an integrated model of prevention focused (proactive vs. reactive) research (Coie et al., 1993; Stormont, Reinke, & Herman, 2010). The basic concept behind the model is identifying malleable factors that give rise to increased risk of long-term negative outcomes, as well as protective factors that are likely to improve long-term outcomes, and then implementing interventions to modify those factors and reduce risk while promoting success (Coie et al., 1993; Lembke, McMaster, & Stecker, 2010; Stormont et al., 2010). Intervention is conceptualized as occurring in levels; the primary

level for the general population, a secondary level targeting elevated risk sub-groups, and a tertiary level targeting high-risk individuals (Coie, et al., 1993; Stormont et al., 2010).

The prevention science model has manifested in education primarily as Response to Intervention (RTI) for academics and Positive Behavior Intervention Supports (PBIS) for behavior in schools (Lembke, et al., 2010; Stormont et al., 2010; Sugai & Horner, 2006). Within the prevention science model, it is important to begin by clearly identifying a target sub-population, the long-term negative outcomes to ameliorate, the malleable risk and protective factors influencing the outcomes in question, and develop tools to identify these risk factors and monitor them across time as early as possible (Lembke et al., 2010; Stormont et al., 2010).

CHAPTER 2

REVIEW OF THE LITERATURE

With the Prevention Science framework in mind, the purpose of the following literature review is to clearly identify the sub-population of interest (ELs), the outcome of interest (writing), and describe the unique features of writing development for ELs that may inform screening for risk and monitoring progress in response to instruction across time.

The Current State and Characteristics of Education for ELs

In 2009, 21% of school age children spoke a language other than English in the home as their primary language and these culturally and linguistically diverse students are increasing in numbers (Aud et al., 2012; Genesee, Lindholm-Leary, Saunders, & Christian, 2005; Soto, Ariel, Hooker, and Batalova, 2015). The Latinx population alone has experienced an estimated 39% increase across the last ten years (Aud et al., 2012). Of these students, those officially identified as EL represent between 9-10% of the overall K-12 student population, about 17% of the 1st grade population, and are the fastest growing sub-population of students (Aud et al., 2012; Kena et al., 2016; NCES, 2015; Soto et al., 2015). In several districts across the U.S., ELs actually represent the student majority (Soto, et al., 2015). Although the majority of ELs live in five states (i.e., California, Texas, Florida, New York, and Illinois), recent census and migration data indicate that ELs and their families are increasingly moving out of urban areas and into less traditional EL states and districts (NCES, 2015; Soto et al., 2015). Kansas, for example, recently had a larger number of new ELs than any other state despite not being a traditional EL state (LeRoy & Flemming, 1939; NCES, 2015; Soto et al., 2015). These

data make it clear that ELs represent a significant segment of the overall student population and are increasingly present across school districts. Therefore, effective identification of ELs in need of academic support(s) is a concern for all districts across all states, especially because ELs are considered to be at high risk for dropping out and academic underperformance (Fry, 2010; Passel, Cohn, & Lopez 2011).

ELs have performed lower than non-ELs in reading and mathematics for all available assessment years beginning in the late 1990s until the most recent available data for the 2014/2015 academic year (Kena et al., 2016). Specific to writing, the 2011 National Assessment of Educational Progress results reported that 99% of all ELs in the 12th grade performed below proficient in writing and 80% of all ELs in the 12th grade failed to perform at even the basic level of writing proficiency. These numbers were almost identical for 8th grade ELs. In fact, ELs performed worse than any other sub-population of students, including those with disabilities. Results from previous years were very similar, indicating no significant changes in performance (NCES, 2012). ELs are not currently meeting proficiency nor has their performance improved over time on state standardized assessments in reading, mathematics, or writing, and they are at increased risk for dropping out of school and not attaining a high-school diploma (Fry, 2010; NCES, 2012; Passel, Coh, & Lopez, 2011).

Key Variables in an EL's Literacy Development

Model of Instruction. All states began assessing English Language Proficiency (ELP) for ELs after *Lau v. Nichols* in 1974. However, the No Child Left Behind Act (NCLB, 2001) mandated that schools report ELP for ELs across the domains of listening, speaking, reading, and writing and that state ELP standards be aligned with state

academic standards (National Academies of Sciences, Engineering, & Medicine (NASEM), 2017). Under this legislation, states were required to identify and report data on ELs, set annual academic objectives for ELs, administer an ELP assessment annually, report the percentage of students meeting proficiency, and track students who have tested out of the school's English as Second or Other Language (ESOL) program for two years (referred to as monitored students). After NCLB (2001), one consortium of states—World Class Instructional Design and Assessment (WIDA)—was formed to create ELP assessments. ELs are often classified as beginning, intermediate, advanced, or monitored (Fox & Fairburn, 2011). Generally, states classify their students based primarily upon how they perform on the ELP test(s) but states vary in how they classify and reclassify ELs (Linquanti & Cook, 2015; NASEM, 2017). Once identified, various models of instruction are endorsed by states and employed by districts (Burr et al, 2015).

Bi-lingual education is when the student is provided academic instruction or support in their native language (L1) while also receiving instruction in English. Bi-lingual instruction can be short term (e.g., students are provided one year or less of support in their L1), long-term (e.g., student receives on-going support in their L1 for the duration of their time in school), or anything in-between. The majority of bi-lingual models function with the intention of withdrawing L1 support and moving the student into English-only at some point (De Jong, 2011).

The current trend across states and districts is the implementation of English-only models of instruction (De Jong, 2011). English-only is when students are provided no instructional support in their L1 and all academic content as well as language instruction is delivered in English (De Jong, 2011; Wiese & Garcia, 1998). Increasing immigration

rates and economic recession have fed the popularity of English-only policies at the state level in general (De Jong, 2011; Ovando, 2003), but NCLB (2001) indirectly promoted English-only and short-term bi-lingual programs in schools by mandating schools report ELP and academic progress for ELs as a specific sub-population (Crawford, 2004; De Jong, 2011; Evans & Hornberger, 2005). Similar to bi-lingual models, there is a spectrum of English-only models of instruction ranging from being placed in general education classrooms with no support to the majority of the day spent in a separate class with a trained teacher of ESOL (TESOL).

ELs in long-term bi-lingual and dual-language immersion models demonstrate initial lags in development of English proficiency during the early elementary grades when compared to ELs in English-only programs (Gennesse et al., 2005), but longitudinal evidence and research with older ELs indicate that ELs taught in long-term bi-lingual models outperform those taught in English-only models across all academic content areas and languages, including in English proficiency, by the later grades (Genesse et al., 2005; Rolstad, Mahoney, & Glass, 2005; Slavin & Cheung, 2005). This indicates that the learning trajectories of ELs vary in relation to the extent to which support and on-going instruction is provided in their L1. English-only models are the most common across the U.S. and are likely to be used with increasing numbers of ELs as they move into more districts with less experience and resources related to teaching ELs. Thus, understanding how ELs develop writing skills within English-only models and developing technically adequate writing assessments for those taught in English-only models will have the most relevance for the largest percentage of ELs.

Oral English Development. Language and literacy are highly interrelated in that development in one domain supports development in the other. According to Abedi and Gandara (2006), "language factors have a greater impact on ELL student performance than any other factors" (p. 43). Generally, ELs take 3 to 5 years to achieve advanced proficiency in oral English, characterized by rapid early progress until the middle ranges of proficiency and slowed progress from there (Genesse et al., 2005). Other researchers have indicated that it takes longer, from 5-10 years, for ELs to become proficient in *academic* English (i.e., English typically used in academic settings, Snow & Uccelli, 2009) and exposure to formal education in the student's L1 can decrease the time needed to meet proficiency in academic English (Abedi & Gandara, 2006; Cummins, 1984; Slama, 2012). Oral English proficiency is closely tied to the development of English literacy skills and ELs generally progress quickly across the beginning stages of literacy acquisition and make much slower progress across the intermediate and advanced stages, in a delayed parallel of the stages of oral English acquisition (August, McCardle, & Shanahan, 2014; Cumming, 2016; Gennesse et al., 2005). In other words, a student will typically progress to the next stage in oral English before progressing to the next stage in English literacy.

ELs appear to use specific strategies coinciding with their level of oral English proficiency, beginning with receptive/memorization strategies (beginning proficiency), moving to interactive/attention seeking strategies (intermediate proficiency), and finally to monitoring of their own language and communication (advanced proficiency; Genesse et al., 2005). The receptive stage is characterized by heavy repetition and memorization. The interaction stage is characterized by sustaining verbal interactions and verbal

attention getting. The final stage is related to question asking and monitoring and repairing communication, essentially metacognition in language use. Effective writers employ self-regulation strategies to plan for writing, monitor their writing, and edit/revise their writing (Graham & Harris, 2009). These stages discussed here broadly parallel the ELP levels used by many schools to classify ELs (e.g., beginning, intermediate, advanced). ELs do need a basic level of oral English knowledge to develop literacy skills in English, but the need for oral English fluency is less for ELs with well-developed literacy skills in their L1, suggesting skill transfer across languages (August et al., 2014; Danzak, 2011). Therefore, not only will an EL's L1 literacy knowledge influence their English literacy performance but the actual nature of their L1 may also influence their learning.

Writing Development. Research indicates that early identification of risk and intervention can prevent later failure for students who are at-risk or with disabilities, including in writing (Berninger & Amtmann, 2003; Graham, et al., 2001). Unfortunately, ELs "are less likely than non-DLLs/ELs to be referred to early intervention and early special education, which may have serious consequences" (NASEM, 2017, p. 10-17). Writing is the primary way in which teachers assess knowledge across academic content areas and often determines a student's ability to access as well as have success in higher education, including vocational school (Graham & Perin, 2007; National Writing Project & Nagin, 2006). Writing has also been identified as a key variable for employers when it comes to hiring, retention, and job promotion (National Commission on Writing, 2003; National Writing Project & Nagin, 2006). However, in order to identify risk early and provide appropriate interventions, valid and reliable assessment is necessary. The

development and selection of appropriate assessment should be based upon an understanding of writing development.

Using the Simple View of Writing as a model for early writing development (Berning & Amtmann, 2003), writing is a working memory and language process that depends upon the simultaneous coordination of transcription skills (e.g., transforming orthographical, phonological, and morphological knowledge into text), text generation skills (e.g., forming the words, sentence(s), and discourse), and self-regulation skills (e.g., employing the executive function and metacognitive skills to plan, monitor, review, and revise text) to produce composition. Writing is also a cultural tool (Ferdman, 1990; Fitzgerald & Amendum, 2007) that conveys an individual's social identity as well as actively creates that identity (Danzak & Silliman, 2014; Ferdman, 1990). Writing develops in English for ELs very similarly in many ways to its development with native English speakers, so cognitive models of the writing process are applicable to both native speaking writers and those writing in a second or other language (Genesee et al., 2005). An exception to the application of the Simple View of Writing to ELs is that both L1 and L2 proficiency influence English literacy development (Babayiğit, 2013; Genesse et al., 2005). Skills in the EL's L1 can transfer over to support English writing development but may also provide points of frustration when the two writing conventions are not in agreement (Dressler & Kamil, 2006; NASEM, 2017). Evidence regarding language transfer supports assessing an EL in both English and their L1 in order to inform instruction as well as employing knowledge of the EL's L1 to anticipate difficulties and teach to analogues across the writing conventions. However, given the sheer number of native languages spoken by ELs, it is often not feasible to assess the student in both

English and their L1. Thus, few states assess an EL's L1 fluency. Therefore, the most effective assessments are those that retain their technical adequacy across a variety of L1s.

There are bidirectional relations between oral language, reading, and writing (Dockrell & Arfe, 2014; Connelly, 2014; Silliman, 2014), so none of the domains are truly independent of the others (Cummins, 1984). Thus, an EL's oral English exposure and proficiency is likely to influence their current performance as well as expected rate of growth in both reading and writing. Writing, however, is often viewed as the more complex of the literacy skills and generally the last to fully mature (Boscolo, 2014; Cummins, 1984; Snow & Uccelli, 2009). Therefore, it should be expected that writing would lag in comparison to oral English and reading development. Differences in writing quality between ELs and non-ELs exist even when controlling for socio-economic status, years of formal English schooling, and spelling; but oral English proficiency remains a critical variable (Babayiğit, 2013). However, transcription level processes (e.g., spelling) explain the most variance in terms of overall writing quality across oral English proficiencies for ELs in the early elementary grades (Harrison et al., 2016).

Chenoweth and Hayes (2001) note that ELs likely spend much of their cognitive resources on the transcription and text generation levels, especially when mentally translating from their L1, leaving fewer resources to spend on self-regulation which result in less planning and reviewing for writing (Chenoweth & Hayes, 2001; Silva, 1993). EL writing falls below that of their non-EL peers more often and more dramatically when coursework demands self-regulation strategies in order to promote variety and complexity in vocabulary and syntax as well as the use of various genres of writing,

especially informational genres (Bulte & Housen, 2014; Campbell, Espin, & McMaster, 2013; Cumming, 2016; Harrison et al., 2016). While spelling, grammar, and other basic mechanical features of writing are critical to success in elementary school, it appears that an EL's growth in syntactic and lexical complexity as well as their ability to change genres, which are all influenced by self-regulation, may be more indicative of long-term success in writing (Bulte & Houssen, 2014; Cumming, 2016; Harrison et al., 2016). Therefore, assessment of young EL fluency in transcription and text generation skills should provide a good indication of whether or not the EL is thinking of English words and sentences and then transcribing them with minimal cognitive effort in order to free cognitive resources for self-regulation. Currently, young ELs appear to slip through the cracks in writing during the early elementary grades only to be identified as a struggling writer after they have developed a long history of failure (Artiles et al., 2005; Klingner, Artiles, & Barletta, 2006). This delay in the identification of risk may have particularly damaging impacts upon long-term writing performance because failure decreases motivation and motivation has been identified as a key variable for successful writers (Bruning & Kauffman, 2016; Graham, Berninger, and Fan, 2007).

Motivation and Writing. Motivation is a broad field encompassing many theories. Defined broadly, motivation comprises the cognitions that energize behavior. Thus, the presence of certain behaviors indicates motivation. Boscolo and Gelati (2006) describe motivation in terms of two inter-related constructs. One construct is the student's sense of competence or ability to perform the task and the other is the student's feeling that the task is relevant, important, or of interest to the student (e.g., values). These two constructs work together to energize specific behaviors, such as task

engagement, task persistence, task completion, and receptiveness to feedback, which are collectively referred to as motivated academic behavior for the remainder of this study. Specific to writing, Graham and Harris (2009) lamented that young writers, especially struggling writers, often have seemingly intractable negative attitudes toward writing as early as the 3rd grade. An EL's motivation for learning English writing may be further impacted by the student's feelings toward learning English in school in general, especially as a cultural practice (Ferdman, 1990; Matutue-Bianchi, 1986). Because ELs have such varied educational histories as well as diverse histories and exposure to the English language, it seems to reason that measures of motivation in writing may be indicative of a student's potential to take advantage of future opportunities to learn in a way that takes into account not only one's feelings of efficacy and value of writing, but also of one's valuation of English as a cultural practice.

DiPerna and Elliott (2002) defined academic enablers as "attitudes and behaviors that allow a student to participate in, and ultimately benefit from instruction in the classroom" (p. 294). One key domain of academic enablers is motivation (DiPerna, Volpe, & Elliott, 2002). According to Pajares (2003), "research findings have consistently shown that writing self-efficacy beliefs and writing performance are related" (p. 144). Graham, Berninger, and Fan (2007) explored a component of interest called attitude toward writing and found that for students in the early elementary grades, attitude toward writing directly influenced a student's writing achievement, underscoring the importance that the construct of motivation plays in writing achievement. Furthermore, studies have shown that self-efficacy and interest are malleable constructs that can be changed in response to targeted instruction (Schunk & Swartz, 1993; Schunk &

Zimmerman, 2007; Zimmerman & Kitsantas, 1999, 2007). Thus, a highly motivated learner is more likely to take advantage of classroom instruction, practice writing more often, and improve at a greater rate than an unmotivated learner. This study will use the phrase 'motivated academic behavior' as a proxy measure of motivation in writing. Motivated academic behavior is defined here as being on-task during writing instruction/assignments, persistence in writing as measured by assignment completion, being receptive to and seeking feedback, a willingness to share written work, and using free time to practice writing. This definition is in-line with that for academic enablers as put forth by DiPerna and Elliott (2002).

Early Identification of Risk and Disability with ELs.

Burr and colleagues (2015) state, "no proven method exists for identifying an English learner student who has a disability and then placing that student in the most appropriate instructional program" (p.1). Further compounding the issue is that states promote a variety of instructional models for ELs, providing little consistency across states and sometimes even within states in terms of how ELs are taught, how they should be identified as EL, how to measure ELP, or how they should be identified as at-risk or with a disability (Burr et al., 2015; Klingner et al., 2006). Not surprisingly, these inconsistencies in identification and service delivery have led to mixed findings regarding the issue of over/under-representation of ELs in special education (Burr et al., 2015; Klingner et al., 2006).

Identification patterns are further complicated by ambiguities and issues inherently related to special education identification itself, especially the categories of Specific Learning Disability (SLD), Mild Intellectual Disability (MID), and Emotional

Disability (ED), which make up the majority of students receiving special education services (Burr et al., 2015; Klingner et al., 2005). For example, African Americans are generally cited as being over-identified for ED and MID (Sullivan & Bal, 2013) but at least one recent study suggested they are actually under-identified (Morgan et al., 2015). These contradictory findings related to over /under identification for special education, even without the additional layer of being an EL, should serve to underscore the tentative nature by which over/under-representation data for ELs should be interpreted. However, when special education identification data are disaggregated by state, district, school, and grade, some patterns do emerge (Klingner et al., 2005).

Artiles and colleagues (2005) found that ELs most at-risk had low proficiency in both English and their L1, patterns of over-representation for all ELs emerged in upper elementary and secondary grades, and ELs in English-only models were more likely to be referred to special education and identified as having a disability. ELs are not likely to be identified as having a disability and receive special education services until they reach upper-elementary or secondary grades, at which point they are more likely to be over-identified (Artiles et al., 2005; Klinger et al., 2006; Rueda, Artiles, Salazar, & Higareda, 2002; Samson & Lesaux, 2008; Sullivan, 2011). Considering that states are increasingly adopting English-only models and the EL population is growing, this is of great concern to the field of special education. The initial under-representation in the early elementary grades seen for ELs likely results from many teachers' and administrators' uncertainties over whether or not insufficient academic progress is a result of a disability or normal patterns of language acquisition and acculturation as well as a lack of services for ELs with disabilities even if they were identified (Burr et al., 2015; Klingner et al., 2005;

Samson & Lesaux, 2008). There appears to be an issue of both under-referral for special education and then over-qualification for those referred, resulting from a poor understanding of the needs and characteristics of EL learners in conjunction with a lack of assessment tools that are properly validated for EL populations (Abedi, 2006; Burr et al., 2015; Sullivan, 2011).

Further compounding identification of risk and disability for ELs are issues related to their representation in the literature and concerns related to assessment in general. Research with ELs often treats them as one homogenous sub-population, but they are actually quite diverse (Artiles, et al., 2005; Klingner, et al., 2006). Despite about 79% of the EL population being Spanish speakers, there is incredible diversity both within the Spanish-speaking majority as well as across the estimated 400 plus other native languages spoken (Kindler, 2002; Passel, et al., 2011; Sandberg & Reschly, 2010). Beyond the many dialects of Spanish and the multitude of other native languages, ELs are also very diverse in terms of socio-economic status, L1 proficiency, years receiving formal ESOL services, ethnicity, migrant status, and cultural background (Artiles et al., 2005; Ferdman, 1990). The use of standardized assessments with ELs in general has been questioned on many grounds but one major criticism is that they are often not representative of the EL population because they are typically presented as one homogenous sub-population, if represented at all (Abedi & Garanda, 2006; Danzak, 2011; Slama, 2012; Wolf, Farnsworth, & Herman, 2008). Abedi and Garanda (2006) suggest using assessment created and/or normed specifically with the EL population and determining achievement and progress based upon that of other ELs with similar initial English proficiency.

Multi-Tiered Systems of Support. Since the reauthorization of IDEA in 2004, many elementary schools have adopted some form of RTI (Linan-Thompson, Cirino, & Vaughn, 2007). RTI involves using screening assessments to identify students who are at-risk for failure and then monitoring at-risk students' learning using regular progress monitoring assessments (Deno et al., 2009; Fuchs, D. & Fuchs, L., 2006). A combination of risk, as measured by the screeners, and the student's rate of learning in response to instruction are used to determine the level of intervention for a student (Burr et al., 2015; Fuchs, D., Fuchs, L., & Compton, 2012). Progress monitoring assessments produce quantifiable data that are technically adequate and sensitive to changes in learning in order to inform instruction and placement decisions within an RTI model (Deno et al., 2009). Additionally, because this assessment is conducted on a regular basis (e.g., weekly or bi-weekly), progress monitoring assessments should be quick and easy to score, administer, and cheap for schools (Deno et al., 2009). CBM are commonly used by schools implementing RTI for both screening and progress monitoring (Fuchs, et al., 2012) and this is fitting given the definition of CBM is a global indicator of academic achievement that is valid, reliable, sensitive to change, and quick and easy to administer and score (Deno, 1985; McMaster & Espin, 2007).

The use of CBM is a central feature to most RTI models and is generally well established for ELs in reading (Keller-Margulis, Payan, & Booth, 2012; Keller-Margulis, Payan, Jaspers, & Brewton, 2016), including published norms for ELs (Pearson, 2012). However, seasonal growth evidenced by non-ELs does not extend to ELs, and initial performance, as well as L1 background, can impact English reading CBM validity and reliability (Keller-Margulis, Clemens, Im, Kwok, & Booth, 2012; Keller-Margulis et al.,

2016; Logan & Petscher, 2010; Betts, Bolt, Decker, Muyskens, & Marston, 2009). Studies have explored reading CBMs in Spanish (Keller-Margulis, et al., 2012), Hebrew (Kaminitz-Berkooz & Shapiro, 2005), and Arabic (Abu-Hamour, 2013; Abu-Hamour, Al-Hmouz, & Kenana, 2013; Abu-Hamour, 2014). These studies have returned promising results for screening and progress monitoring reading for young ELs within an RTI model but the research base behind CBM-Ws for ELs is still nascent. CBM will have increasing relevance for ELs because ESSA encourages states to go beyond annual summative assessment results to include benchmark assessments measuring growth, such as CBM (NASEM, 2017). Furthermore, the NASEM highlighted the use of progress monitoring assessment as an effective instructional practice for young ELs (2017).

Current Models of Writing Assessment for ELs. Few assessment tools have been researched with ELs specifically (Abedi & Garanda, 2016; NCRI, 2011). WIDA (2013) recommend using a number of authentic assessment procedures emphasizing assessment for learning, including; observations, collection and evaluation of student work, achievement assessments, conferencing with students, rubrics, rating scales, portfolios, and field notes. This information is indeed what classroom teachers should and often do collect as a part of sound pedagogy and should be used to inform placement decisions (Burr et al., 2015), but it is not likely to be collated and analyzed on a weekly or bi-weekly basis as a type of formalized progress monitoring and does not easily allow for objective interpretation for identification of risk and effectiveness of instruction within an RTI model. Moreover, no studies have been found exploring the effectiveness of classroom observations and other non-test data in identifying disability for ELs (Burr et al., 2015).

Curriculum Based Measures. Data driven instruction and data-based decision making (DBDM) are critical to any RTI model, requiring quantified data. In order to be effective, DBDM requires assessments providing quantifiable data, such as CBM (Deno et al., 2009). The majority of recommended writing assessments for ELs are holistic in nature (Murphy, 2009; WIDA, 2014), do not lend themselves well to regular progress monitoring, and can be vulnerable to teacher bias (Rezaei & Lovon, 2010). CBM-W data collected in conjunction with data recommended by WIDA (2014) could help promote data-based decisions, improve timely access to special education services, and provide a better picture of the student's overall development and growth in writing. However, according to Fuchs (2004), 3 stages of research need to occur in order to establish the utility of CBM; including 1) establishing the technical features of static scores, 2) examining technical features of slope, and 3) examining instructional utility. There is a growing body of research providing support for CBM-Ws with non-ELs (Romig, Therrien, & Lloyd, 2017). Specific to young writers, three common forms of CBM-W are Story Prompt (SP), PW, and WD. A brief review of the literature regarding each form of CBM-W is provided in the following sections.

Word Dictation. WD captures growth in word level transcription skills according the Simple View of Writing but has also been evidenced to have good predictive validity with more comprehensive writing outcome measures such as the Test of Early Written Language-II (TEWL-II, Hresko, Herron, & Peak, 1996; Lembke, Deno, & Hall, 2003). Specifically, WD has demonstrated evidence of technical adequacy for use in screening and progress monitoring in grades 1-3 (reliability: $r > .70$, validity: $r > .50$; Ritchey & Coker, 2013; Hampton & Lembke, 2016; Lembke et al., 2003). Additionally, word

spelling was accurate at identifying writing risk in first grade (AUC = .780-.873; Ritchey & Coker, 2014). For WD, the test administrator reads as many words as the student can write in 3 minutes, up to 30 words. Each word is only read two times. WD is scored using a variety of approaches that are best categorized as simple production, accurate-production, and production independent.

Simple production scoring procedures for WD include words written (WW). WW is calculated by adding the number of words the student attempted. Accurate-production scoring procedures include words spelled correctly (WSC; counting the number of words the student spelled correctly) and correct letter sequences (CLS). CLS involves counting each sequence of letters as either correct or incorrect, including beginning with the correct letter and ending with the correct letter. For example, if the word provided was 'fire' then this word would have 5 potential CLS but if the student wrote 'fir' then they would only be given credit for 3 CLS and would also be given 1 incorrect letter sequence (ILS). Another accurate-production metric that has been explored is correct minus incorrect letter sequences (C-ILS). However, this metric can produce negative scores for many low-performers. Production independent scoring procedures include %WSC (WSC divided by WW) and %CLS (CLS/ (CLS+ILS)). Production independent measures are prone to ceiling effects wherein a student may only attempt one word yet spell it correctly, resulting in 100% CLS or 100% WSC.

Picture Word. PW is another promising measure that captures growth in sentence writing skills at both the text generation and transcription levels according to the Simple View of Writing (Lembke et al., 2003; McMaster et al., 2009; McMaster, Du et al., 2011). PW has evidence of technical adequacy for students in the 1st through 3rd grades,

including sensitivity to growth (reliability: $r > .70$, validity: $r > .50$; McMaster et al., 2011; McMaster et al., 2009). PW consists of 12 pictures with a key word below each picture and students are asked to write their best sentence for as many pictures as possible in 3 minutes (McMaster et al., 2014). PW can be group administered rather than requiring individual administration, as WD does (McMaster et al., 2014). Therefore, PW may be a more feasible universal screener and tool for progress monitoring for practicing teachers. As is WD, PW is scored in a variety of ways best categorized as simple production, accurate production, and production independent.

Similar to WD, a simple production metric of WW for PW entails counting the number of words the student wrote whether they were correct or not. Accurate production measures include WSC (counting the number of words written that are spelled in a way that accurately makes a real word, regardless of context) and Correct Word Sequences (CWS). CWS entails scoring whether or not the word written is spelled correctly and makes sense within the context of the sentence, inclusive of grammar, capitalization, and punctuation. For example, if the student were to write, "I lick dogs because they fetch" then the student would receive a CWS for starting the sentence with a capital 'I', but no CWS from 'I' to 'lick' or 'lick' to 'dogs' because that does not make sense within the context of the sentence. The student would receive CWS for 'dogs' to 'because', 'because' to 'they', and 'they' to 'fetch'. However, the student would not get a CWS after 'fetch' because they failed to use end punctuation. Thus, the student would receive 4 CWS. An Incorrect Word Sequence (IWS) is scored when a word is incorrect. In the previous example, the student would have received 3 IWS. Another accurate-production measure that has been explored is CWS minus IWS (C-IWS). Production independent

metrics include %WSC (WSC/WW) and %CWS (CWS/(CWS+IWS)). The last three measures have the same faults as C-ILS, %WSC, and %CLS for WD.

Story Prompt. SP consists of providing the student a story starter, such as, "One day I came to school but no one was there, so I…". The student is then given a set amount of time to think about and plan their story and a set amount of time to write their story. SP is scored using the same metrics as PW, including: WW, WSC, CWS, and IWS. The times provided to think and actually compose stories vary according to who produced the form and the age of the student being assessed. However, young students are generally provided 30 seconds to 1 minute to plan and 3 minutes to compose their story. SP has demonstrated evidence of test-retest reliability (r = .64-.70) and criterion validity ($r \geq .70$) with standardized tests of writing for students in grades 3-6 (Deno, Mirkin, & Marston, 1980; Marston & Deno, 1981) and in grades 1-3 (reliability: $r > .70$; McMaster & Campbell, 2008; McMaster, Du, & Petursdottir, 2009; McMaster et al., 2011; Ritchey & Coker, 2013). SPs also differentiated between grade levels and were sensitive to growth over an academic year (Deno et al., 1982; McMaster et al., 2011).

ELs and CBM-W. A review of the literature was conducted for studies exploring the technical adequacy of CBM-W with ELs across 6 databases (Google Scholar, ERIC, Education Full-Text, Scopus, Academic Search Complete, & Psych Info). Inclusion criteria were that the article had to be published in an English peer-reviewed journal, the participants had to be K-12 students and identified as EL, and the study had to include a CBM-W. The search returned 4 articles (Campbell, 2010; Campbell et al., 2013; Espin et al., 2008; Keller-Margulis et al., 2016). Two articles (Campbell, 2010; Campbell et al., 2013) came from data gathered from a single dissertation study (Campbell, 2006) and

another (Espin et al., 2008) should be considered exploratory in nature because it was not specifically designed for ELs. Espin and colleagues (2008) designed their benchmarking study for the general population and happened to have a sample of ELs large enough to warrant separate analyses. Three studies included high-school age participants (e.g., Campbell, 2010; Campbell et al., 2013; Espin et al., 2008) and one included 4th graders (e.g., Keller-Margulis et al., 2016). The sample sizes across the studies ranged from N = 36 (Campbell et al., 2013) to N = 57 (Campbell, 2010). Three of the 4 studies were conducted in Minnesota and the majority of ELs in those studies spoke various African/Arabic languages as their L1. The study by Keller-Margulis (2016) included predominately Spanish-speaking ELs and was conducted in the Southwest.

The only study to include participants with beginning English proficiency was Campbell (2010) and used passage-copying CBM-W. Passage-copying CBM-W yielded correlations with the criterion measures (teacher rating scale, Minnesota Basic Standards Test (MBST), Test of Emerging Academic English (TEAE), and Test of Written Language-3 (TOWL-III)) ranging from .38-.67, which is considered low moderate to moderate and in line with previous research employing the same measure with non-ELs (Campbell, 2010; McMaster & Campbell, 2008). The best metric was CWS, correlations ranging from .50-.67. Passage copying is a common practice in ESOL classrooms and the passage copying CBM is very easy to administer and score (Campbell, 2010). Therefore, it shows some promise as a screening assessment for beginning proficiency high school ELs but there is no evidence to support it as a tool for progress monitoring or screening with young ELs. However, only one study examined CBM-W with elementary age ELs.

Keller-Margulis and colleagues (2016) examined the validity and diagnostic accuracy of SP with 19 predominately Spanish-speaking ELs and 31 Spanish speaking monitored students in the 4th grade. The criterion measure was the Texas state writing test (STAAR-W). Results found that only production independent (%CWS) and accurate production (C-IWS) measures captured the complexities of writing similar to the STAAR-W. The metrics with significant correlations across at least two groups (non-EL, EL, monitored) were %CWS, %WSC, and C-IWS. As opposed to Espin et al. (2008), who examined SP with high-school ELs, correlations were stronger for non-ELs than for ELs. Correlations varied across time of year with only winter producing significant results for ELs and only spring producing significant results for monitored students. The metrics showing the most promise for ELs were %CWS and C-IWS. They used Receiver Operating Characteristic (ROC) curve analysis with both sensitivity (i.e., screeners ability to identify risk for those that score as at-risk on a criterion measure) and specificity (i.e., screeners ability to identify as no-risk those that do not score as at-risk on a criterion measure) set at a minimum of .70 and found that it was not possible to find adequate cut scores for the majority of metrics. The Area Under the Curve (AUC) statistic was not significant when cut points were found. This indicates that SP is not a significant predictor of STAAR-W test performance for the participants included in the study (Keller-Margulis et al., 2016). Thus, it seems that SP is not an appropriate CBM-W for ELs in the elementary grades.

Summary and Conclusion

ELs represent a diverse and significant sub-population of students in U.S. public schools and are at increased risk for failure in writing (Genesee et al., 2005; NAESM,

2017). Early identification of risk and intervention in writing can reduce risk and support future academic performance (Berninger & Amtmann, 2003; Graham, Harris, & Larsen, 2001) but ELs are less likely to be identified as at-risk and receive early intervention supports (NAESM, 2017). In order to promote early identification of risk and early intervention, reliable and valid assessments are needed for screening and progress monitoring (Burr et al., 2015). Few assessments have been validated specifically for ELs, especially in writing (Genesee et al., 2005). Oral English proficiency, L1 literacy proficiency, and model of ESOL instruction are key variables influencing the learning trajectories of ELs (Genesse et al., 2005; NAESM, 2017). However, English-only is the primary model of instruction for ELs and is likely to increase as ELs move to areas with less experience and resources to support them. Although L1 literacy proficiency is a key variable, the sheer volume of native languages spoken by ELs often precludes the collection of such data, therefore assessments that have been validated across a multitude of native languages have the most utility. Specific to early writing, fluency in transcription and basic text generation skills are critical in freeing up the cognitive resources needed to support self-regulation as ELs compose longer and more complex compositions in the upper-elementary grades.

WD and PW assess early writing in the domains of transcription and text generation, respectively. Thus, WD and PW show promise for assessing the overall writing skills of young ELs. However, ELs represent a unique sub-population and initial risk and rates of weekly growth should be based upon the performance of other ELs taught within similar models of instruction and with similar initial oral English proficiencies (NAESM, 2017). Furthermore, ELs come to U.S. classrooms with diverse

prior academic and language experiences in both English and their L1, possibly impacting the validity of static CBM-W scores. Although parallel assessment should be conducted in the EL's L1, this is often not feasible due to the multitude of languages spoken by ELs. Additionally, it is often difficult to readily accumulate data and quantify prior opportunities to learn, in both English and the EL's L1. In lieu of quality data quantifying an EL's prior experiences, the incorporation of motivated academic behavior may be indicative of an EL's likelihood of maximizing future opportunities to learn and therefore improve the predictive abilities of CBM-W. This study will explore the technical adequacy of WD and PW for ELs speaking a variety of native languages and taught in English-only models in the 1st through 3rd grades. Areas of technical adequacy to be explored include reliability, predictive validity, concurrent validity, divergent validity, the impact of oral English proficiency upon CBM-W performance, and the sensitivity to growth for each measure. Furthermore, researchers caution against using benchmarks established with the general population for identifying risk and generating goals for ELs and this study will explore EL benchmarks in comparison to the general population. Additional analyses will explore the impact of integrating a measure of observable motivated academic behavior with WD and PW upon predictive validity and diagnostic accuracy.

CHAPTER 3

METHODS

The purpose of this study was to examine the technical adequacy of two forms of CBM-W, WD and PW, with ELs in the 1st-3rd grades. Specifically, the study examined the inter-scorer reliability, alternate form reliability, and criterion validity of WD and PW at three time-points (fall, winter, spring) using two criterion measures: (1) ACCESS-W and (2) MAP-ELA. The study also explored the sensitivity to growth over time of WD and PW for ELs. Additionally, overall performance of the general population on WD and PW were descriptively compared to that of the ELs in this study. Finally, this study examined the impact of a measure of motivated academic behavior (Social, Academic, and Emotional Behavior Risk Screener, SAEBRS; Kilgus, Chafouleas, & Riley-Tillman, 2013) in writing on the predictive capabilities and diagnostic accuracy of both PW and WD.

Research Questions

1. What is the technical adequacy of WD and PW as indicators of general writing performance for ELs in the 1st-3rd grades?
    a. What is the inter-scorer and alternate form reliability of WD and PW?
    b. What is the concurrent and predictive validity of WD and PW with the ACCESS writing sub-test?
    c. What is the sensitivity to growth for WD and PW for English language learners?
    d. How do WD and PW performance by ELs compare to the general population?

2. How does the inclusion of a measure of motivated academic behavior in writing impact the predictive validity and diagnostic accuracy of WD and PW for ELs?

Participants and Setting

Sampling. Three districts were originally recruited for participation in this project. Two were in large urban areas, one in the Northeast and the other in the Midwest, while the third was in a mid-size city in the Midwest. By the time IRB approval was received and consent letters were prepared, all but the district in the mid-size city in the Midwest had dropped out of the study. The primary reason provided for withdrawing from the study was time lost to instruction for assessment. Elementary ESOL teachers were the next line of contact and teacher consent letters were sent to 17 teachers, of which 10 consented to participate and participated throughout the study, for a 59% return rate by ESOL teachers. Letters were sent to ESOL teachers in 15 school buildings and letters consenting to participate were returned by teachers by at least one ESOL teacher in 9 schools, for a school participation rate of 60%. Several teachers stated that they would not participate because the district was planning to relocate them to different schools at some point during the academic year but most provided no reason. The estimated total number of EL students in grades 1st-3rd across the district was 380, of which parental consent letters were sent home to approximately 230. Seventy-three students returned parental consent, signed student assent to participate in the study, and completed the study for a participation rate of 32% of those receiving parental consent letters and 19% across all 1st-3rd grade ELs in the district. One ESOL teacher initially consented to participate but withdrew from the study during the fall benchmarking period; however, this teacher was not counted as one of the 10 participating teachers. Furthermore, only

one student consented to participate for the teacher that withdrew from the study and the student was not included as one of the 73 student participants.

School and District Information. All student participants received ESOL services from one of 10 ESOL teachers across 9 elementary schools from a single school district in the Midwest. The district is a mid sized school district serving 18,000 K-12 students during the 2016-2017 academic year. Across the district, students were 60.8% White, 20% Black, 6.3% Hispanic, and 5.5% Asian. Furthermore, 39.7% of students were eligible for free and reduced price lunch, 9.7% had IEPs, and 6% received ESOL services during the 2016-2017 academic year. The demographics of the specific schools that participants attended in this study were 52.1% free and reduced priced lunch, 9.8% with IEPs, and 12% receiving ESOL services. Demographics are provided in Table 1 for each school. All nine teachers were female ESOL teachers. ESOL services were provided using a variety of models across schools ranging from primarily co-teaching/push-in models to pullout models for ESOL instruction. However, the district followed English-only practices and no bi-lingual or dual-language instruction was implemented within any of the schools. Thus, the common model of instruction across all participants was English-only ESOL services.

Table 1

Participant School Demographics

School	*Total Pop*	*%FRL*	*%White*	*%Black*	*%His*	*% Asian*	*%EL*	*%IEP*
1	415	83.2	31.3	36.1	20.5	3.4	21	11.8
2	449	100	37.2	37.6	10.9	8.5	18.3	13.6
3	499	30.4	68.7	14.0	4.8	7.8	12.6	5.8

4	411	100	47.5	33.8	7.1	-	4.6	11.0
5	691	29.6	61.1	11.6	5.6	12.5	17.2	8.8
6	240	56.8	56.7	10.4	19.6	3.3	19.2	14.6
7	508	25.8	71.3	11.2	3.4	5.5	5.7	9.5
8	464	28.2	71.6	10.8	4.7	4.5	6.3	8.6
9	379	27.6	78.1	6.6	5.5	2.9	4.2	9.2

Student Participants. The district used the ACCESS test to identify students as qualifying for ESOL services and to determine ELP. The ACCESS test is generally administered in late January each school year and the results are not typically made available to administrators and teachers until the end of the school year. Students arriving to the district after the ACCESS testing window that identify as non-native English speakers take the ACCESS screener called the WIDA-ACCESS Placement Test (W-APT) and are assigned a proficiency level and ESOL services accordingly. All participants were identified ELs receiving ESOL services as either a function of the 2015-2016 ACCESS test scores or W-APT scores. Seventeen student participants had incomplete 2015-2016 ACCESS data or were new arrivals and took the W-APT, representing 23% of the sample, although a general English proficiency level was assigned and provided for each participant. Therefore, each subject's English proficiency scores reported here and used for analysis are drawn from their January 2016-2017 ACCESS test scores.

The total number of participants in this study was 73 ELs across grades 1-3. Table 2 provides sample size by grade and information regarding gender, ELP, and native-language. ELP is described as beginning, intermediate, or advanced. Participants were

considered beginning if their overall ACCESS proficiency score was 2.9 or less, intermediate if their score was 3.0 to 4.9, and advanced if the student scored above a 4.9. As noted in Table 2, only 2 students were advanced, both in 1st grade, with 21 at beginning levels and 48 at intermediate levels. The sample is weighted toward the beginning and intermediate range of ELP. 64.4% of the participants were male. The three most common foreign languages spoken were Spanish (38.4%), Arabic (9.6%), and Burmese (5.5%). Other languages represented were Tigrinya, Korean, Vietnamese, Czech, Karenni, Chinese, Kirundi, Tagalog, Swahili, and unspecified. Two students did not take or had incomplete data on the 2016-2017 ACCESS. Thus, only 71 of the original 73 participants had complete ELP data to report. The racial demographics of the sample were 46% Hispanic, 29% Asian, 12% White, 9% Black, and 4% Other or Multiracial.

Table 2

Student Participant Demographics

Grade	N	%Male	# ELP Beg/Int/Ad	%Spanish	%Arabic	# Other Lang.
1	24	67%	6/15/2*	41.7%	20.8%	6
2	25	60%	7/18/0	28%	4%	9
3	24	67%	8/15/0*	45.8%	4.2%	5
Total	**73**	**64%**	**21/48/2***	**38.4%**	**9.6%**	**11**

Note * = one student each did not complete ACCESS in grades 1 and 3. ELP = English Language Proficiency. Beg = Beginning Proficiency. Int = Intermediate Proficiency. Ad = Advanced Proficiency.

Predictor Variables

CBM-W. All CBM-Ws (WD, PW) were administered and scored according to standard procedures published by McMaster et al. (2014). Each type of CBM-W allocates 3 minutes for the student to write. However, one modification was provided to the

scoring procedures for all CBM-W forms. Students were encouraged to write in English but any responses written in their L1 were not scored nor factored into the scoring metrics. Considering that the most technically adequate measures are those that include accuracy, such as CWS, and that the majority of elementary teachers are mono-lingual, it is not feasible for ESOL teachers to be able to adequately read and score the accuracy of writing in all other languages. Furthermore, if teachers were directed to score writing in the EL's native language according to accuracy then that would negate the standardization and utility of scoring rules if only a minority of teachers are able to properly apply them and only in the one or two languages they may be able to read and score. Additionally, scoring writing produced in the EL's L1 as incorrect or as an error, even if used correctly, is in conflict with research showing that more successful ELs view and use their L1 as a resource for writing (NASEM, 2017) and also works to devalue the student's L1.

Word Dictation. WD-CBM was selected because it is a measure of transcription skills at the word level and transcription skills explain the most differences between L1 and L2 writing for young EL students (Harrison et al., 2016). WD has demonstrated adequate reliability ($r \geq .89$-$.95$), validity ($r = .29$-$.75$), and is an effective measure for weekly or bi-weekly progress monitoring (Hampton et al., 2016; Lembke, et al., 2003; Lembke et al., 2015). WD has also been shown to accurately classify students as at-risk according to the WIAT Spelling subtest for second (AUC = .86) and third (AUC = .87) grades (Lembke et al., 2015). Several versions of WD are available, including guidelines for teachers to create their own (Hosp & Hosp, 2003). However, this study implemented WD forms created by McMaster and colleagues (2014) because they are standard forms

that have technical evidence directly supporting them with 1st-3rd students and the forms use spelling patterns indicated by the Common Core State Standards (Lembke et al., 2014). WD is individually administered for 3 minutes, the administrator dictates each word twice, and the student is to write as many of the 30 words as possible in 3 minutes. In the case that a student completes all 30 words in less than 3 minutes, scores are prorated by calculating a score per second and then multiplying by 180 seconds.

Picture Word. The study used PW forms created by McMaster et al. (2014) because they are standard forms that have been specifically validated with 1st-3rd grade students (Lembke et al., 2014). PW was selected because it is a good measure of sentence level transcription and text generation skills. Furthermore, PW requires the student to generate vocabulary and vocabulary knowledge, which are crucial for ELs. Also, PW may be administered to a whole group and WD must be individually administered, meaning that PW may be a more feasible screening and progress-monitoring tool. PW has adequate reliability and validity for non-ELLs as well as promise as a tool for progress monitoring (Lembke, et al., 2003; McMaster et al., 2009; McMaster et al., 2011). Alternate-form reliabilities for 3-min samples have been reported as $r > .70$, with reliability ranging from $r = .81$-$.84$ in the fall, to $r = .87$-$.91$ in winter, and $r = .86$-$.90$ in spring. Criterion-related validity has ranged from $r = .50$ to $.60$. Weekly prompts have produced reliable slopes within 8 weeks and are sensitive to growth (McMaster, Du, Yeo, Deno, & Ellis, 2011; Lembke et al., 2015). In terms of diagnostic accuracy, PW was most effective for predicting performance on the WIAT Sentence Combining test at first (AUC = .79) and second (AUC = .82) grades (Lembke et al., 2015).

Social, Academic, and Emotional Behavior Risk Screener. The Social, Academic, and Emotional Behavior Risk Screener (SAEBRS) is a 19-item teacher rating scale that has been evidenced to have sufficient reliability, concurrent validity, and diagnostic accuracy (Kilgus, Chafouleas, & Riley-Tillman, 2013; Kilgus, Sims, Nathaniel, & Taylor, 2016; Kilgus, Sims, van der Embse, & Riley-Tilman, 2015). SAEBRS was selected over other screeners based upon its emerging technical adequacy, its brevity, its focus upon both risk and protective behaviors, and the inclusion of an Academic Behavior specific sub-scale. SAEBRS consists of 3 subscales; Academic Behavior (AB; 6 items); Social Behavior (SB; 6 items), and Emotional Behavior (EB; 7 items) and takes 1-3 minutes for a teacher to complete per student. Of specific interest to this study is the AB sub-scale "defined as behaviors that promote (e.g., academic enablers) or limit (e.g., attention problems) one's ability to be prepared for, participate in, and benefit from academic instruction" (Kilgus et al., 2016; p. 22). The definition of the AB sub-scale aligns with this study's definition of motivated academic behavior. AB specific items include: (a) *interest in academic topics,* (b) *preparedness for instruction,* (c) *production of acceptable work,* (d) *difficulty working independently,* (e) *distractedness, and* (f) *academic engagement.*

SAEBRS uses a 4-point Likert scale (*Never, Sometimes, Often, Almost Always*) on which teachers are asked to rate "How frequently the student displayed each of the following behaviors during the previous month." Higher scores on each sub-scale indicate more adaptive functioning. However, motivation and the accompanying motivated academic behaviors is often domain and context specific (e.g., one may display motivated academic behaviors in reading but not math; Bandura, 1997; Bruning &

Kauffman, 2016; Zimmerman & Bandura, 1994) and the SAEBRS is designed to be a screener of general risk (i.e., across content areas). The focus of this study is motivated academic behaviors specific to writing. Therefore, a cover letter was included with the SAEBRS that directed the teacher to, "complete this scale while thinking of the student's behaviors over the past month during *writing instruction* or while *completing writing assignments*".

Criterion Variables

ACCESS. The ACCESS test was used as a criterion variable because it was specifically developed for and normed using ELs (Fox & Fairburn, 2011). Furthermore, performance on the ACCESS test determines whether or not a student receives ESOL services and to some extent the nature of those services, so their performance on this test has real implications and high social validity for ELs. Additionally, ACCESS is the measure by which the district reports EL growth in ELP in accordance with federal accountability laws for NCLB (2001). Thus, the ACCESS test can be considered a high-stakes test for districts that is specific to ELs.

All ELs in WIDA states take the ACCESS in January of every school year, for a total of 35 states plus the District of Columbia (D.C.) and the U.S. Virgin Islands. ACCESS transforms scores into single digit proficiency levels for each sub-test as well as for an overall proficiency score. Scores of 1 or 2 often correspond to beginning, 3 and 4 correspond with intermediate, and above 4.9 represents advanced until the student meets the state's defined cut-off for proficiency. Test scores are also provided as scaled scores. Scaled scores allow for comparison of previous performance across grades and grade clusters. Composite proficiency levels drawn from participant's ACCESS performance in

January of the school year preceding the study were used to classify the student as an EL at the on-set of the study. In the case of new ELs to the district, the W-APT scores were used to classify students as EL. Performance on the ACCESS writing sub-test completed during January of the year during the study will be used as a criterion variable as well as for examining diagnostic accuracy, with performance in the beginning range considered to be 'at-risk' and those performing above the beginning range considered 'no risk'.

With respect to technical adequacy for the ACCESS test, the reading and writing sub-tests are the strongest predictors of performance on content area tests (Paker, Louie, O'Dwyer, 2009). ACCESS sub-tests were correlated with older generations of ELP tests and writing had the second highest average correlation (.561) behind only reading (.765; Yanosky et al., 2013). Writing scaled scores correlated with the listening ACCESS sub-test at .783, the speaking subtest at .575, and the reading sub-test at .895 (Yanosky et al., 2013). Additionally, the ACCESS writing sub-test significantly predicted ($p<.001$) performance on the writing subtest of the New England Common Assessment Program for both 5th and 8th grade ELs (Yanosky et al., 2013).

Time allocated for completion of the writing sub-tests is 35 minutes for 1st graders and 65 minutes for 3rd graders. The writing sub-test is the last sub-test to be administered and is scored by WIDA professionals trained to use their rubric. Scoring by trained WIDA professionals results in a delay between test completion and receipt of results despite ACCESS recently being moved on-line and underscores the purpose of ACCESS as a diagnostic and summative assessment, not a screener or tool for regular progress monitoring. The W-APT takes approximately one hour to administer and is scored by administrators in the school, such as trained ESOL teachers. Scored writing samples are

provided with the W-APT model for scorers to train with. I was unable to locate any technical reports on the W-APT or studies involving the reliability of teacher's use of the rubric. The time needed to administer and score the W-APT raises doubts regarding it's utility as a regular progress-monitoring tool and the lack of available technical adequacy tech reports raises concerns regarding its validity and reliability. Therefore, W-APT results and ACCESS score from the year preceding the study were only used to identify students as ELs to participate in this study. Only results from the 2016/2017 ACCESS were used to determine actual ELP.

MAP. The MAP is a state assessment administered on-line every spring for the purpose of reporting as outlined in NCLB (2006) and detailed technical evidence can be found at: https://dese.mo.gov/sites/default/files/asmt-gl-2016-tech-report-ela-and-math.pdf (Missouri Department of Elementary and Secondary Education (DESE), 2016). MAP includes multiple choice, technology-enhanced, evidence-based selected responses, and short answers. Writing prompts are only administered in grades 5-8. MAP-English Language Arts (MAP-ELA) and Mathematics (MAP-MA) are given to students in grades 3 through 8 and a Science sub-test is added for grades 5 through 8. In 2016, 560 districts and charter schools administered MAP-ELA and MA and 2017 represents the 12th year of administration in the state of Missouri, though the tests have gone through several rounds of revision with 2016's test representing a new baseline.

MAP is designed to inform stakeholders regarding student progress toward state standards. Reliability analysis indicated that MAP was relatively reliable, the uni-dimensionality of each sub-test was confirmed via principal component analysis, and correlations confirmed that the different sub-tests were highly but not perfectly related to

each other via correlations. Scaled scores are transformed and reported as one of four achievement levels: *Below Basic, Basic, Proficient,* or *Advanced.* Both MAP-ELA and MA are administered across multiple sessions and take approximately 100 minutes for each subtest, though tests are untimed and time is approximated. For 3rd grade MAP-ELA, response types include selected response (i.e., multiple choice) and technology-enhanced items (i.e., automatically scored items using drag and drop, drop-down menus, and matching). Although the assessment of writing processes is included within the objectives measured by the MAP-ELA for 3rd graders, they are not required to respond to writing prompts.

Procedures

Recruitment/Consent. Several districts were initially contacted in order to gauge interest in the study and develop a feasible plan that met district needs and inform IRB materials. Once IRB approval was granted by the University of Missouri, only one district decided to participate. Final approval from the remaining district was not granted until early October. Once final approval was granted, all district ESOL teachers were contacted via e-mail soliciting their participation in the study. ESOL teachers were asked to complete a SAEBRS form for each participating student in fall, winter, and spring; collect consent forms sent home with students; inform researchers about needed translations for home communication; assist with communication with home and students during assessment; and work with researchers to schedule and arrange for assessment. Eleven teachers initially signed and returned a letter of consent to take part in the study by mid-October but one withdrew during the fall benchmarking period. Consent letters for the parent/guardian of student participants were translated into various languages

according to feedback from the ESOL teachers. Students were also asked to provide written assent by signing or putting an X on the signature line to participate in the study. ESOL teachers were asked to remain present as the study was explained to potential student participants and letters of consent were sent home to help with communication. Signed letters of consent for student participants were not returned until late-October and early-November, explaining why the fall benchmark period did not begin until mid-November.

Assessment Schedule. Students took two alternate forms (A, B) each of WD and PW in the fall (mid-November), winter (late-January), and spring (mid-April). The forms were counterbalanced across students and time points. A total of 51 students (70%) received form A first and form B second in the fall, form B first and form A second in the winter, and form A first and form B second in the spring. The remaining students were administered forms in the opposite order. Both WD and PW were administered on the same day, often in the same session. Order of administration was set at the school level according to projected numbers at the start of the study. Unfortunately, two teachers and several students withdrew from the study during or just after the first administration, which explains why counterbalancing of order of administration was not more equitable. One teacher removed himself from the study because of the time requirement and another removed herself from the study because she changed schools during the fall benchmarking period. Because the teacher was the gatekeeper to student access as well as the main contact for test organization and the person responsible for completing the SAEBRS, all consented students associated with the teachers dropping out of the study were also removed from the study. The remaining 9 consented teachers remained

throughout the study. Again, total number of student participants was 73 but 2 did not complete the ACCESS-W during the 2016/2017 administration.

All CBM-W were administered within a two-week window across teachers and schools for each benchmark. At the beginning of the CBM-W assessment window, teachers were provided a link to the SAEBRS form to complete for each consented student using Qualtrics, an on-line survey service contracted by the University of Missouri. All teachers completed the SAEBRS for each student within two weeks of completing the CBM-W assessments. Additionally, all students completed the winter benchmark assessments within two weeks of completing the writing sub-test of the ACCESS test in winter. Each benchmark was separated by a minimum of 8 weeks. The MAP was only administered to 3rd grade participants during the spring, within 2-3 weeks of spring CBM-W administration.

Data Collection. As stated previously, SAEBRS data was collected via Qualtrics and completed by each student participant's ESOL teacher. A total of 6 research assistants were trained and administered CBM-W to students. Four of the CBM-W administrators, including the author, were highly experienced and familiar with both types of CBM-W after having participated in a large federally funded grant using both forms of CBM-W prior to and during the current study. The other two administrators were trained by the author in CBM-W administration. Training consisted of reading through standard administration directions, a brief overview, and practice administering each type of CBM-W to the lead researcher until 100% fidelity was met. Additionally, each of the newly trained administrators observed a more experienced administrator give each type of CBM-W to student participants. All ACCESS tests were teacher

administered using the ACCESS on-line version. All teachers had received several hours of training prior to test administration and administered according to WIDA and state recommendations. The writing sub-test was sent to and scored by trained WIDA professionals. The MAP was also administered on-line and scored by state trained assessors.

Approximately 10% of all CBM-W administrations per administrator were evaluated for fidelity of administration by a trained observer using a modified version of the Accuracy of Implementation Rating Scales (AIRS; Fuchs et al., 1984). WD was administered individually for a total of 416 administrations, 36 of which were observed and scored using the AIRS for 9% of administrations. Each administrator was observed at least once and fidelity of administration was calculated by dividing the number of observed elements in the AIRS by the number of possible elements. Total fidelity of WD administration was 99%, with only one error in the fall administration in which the administrator forgot to tell the student to draw a line through any mistake rather than erasing. PW, which was often administered in small groups, had 240 total administrations of which 28 were observed for 12% of administrations. Fidelity of PW administration was 99% with the only error being that an administrator failed to read all of the key words prior to administration for the 2nd administration of PW in the spring.

CBM-W Scoring. Four individuals were responsible for scoring all CBM-W data and all four had extensive prior experience with scoring both forms of CBM-W as researchers on a large federally funded grant in early writing and CBM-W. Two of the scorers, author included, were PhD students and had received over 20 hours of training and practice in scoring each type of CBM-W. The other two scorers held advanced

degrees in school psychology or school administration, had over 20 years of experience working in public schools, and had well over 20 hours of training and experience in scoring both forms of CBM-W. The author was the primary scorer of all CBM-W data in fall and winter. A second scorer independently scored a random selection of 24% of all student CBM-W assessments across the 3 time-points (fall, winter, spring). In fall and winter, the author was the primary scorer for all CBM-Ws and a second scorer independently scored a selection of CBM-Ws. The total percentage of CBM-Ws scored at fall and winter for inter-rater reliability were 22% and 26% respectively. For the spring benchmark, three trained scorers served as the primary scorers of all data, inclusive of the author, for a total of 22% scored for inter-rater reliability during spring with a minimum of 20% of forms checked for each primary scorer. Scoring reliability was calculated by dividing total scoring agreements by agreements plus disagreements. Reliability was calculated for WW, WSC, CLS, and ILS for WD as well as WW, WSC, CWS, and IWS for PW. Scoring procedures were followed as published by McMaster et al. (2014) with the exception of not scoring responses written in the student's native language.

Data Preparation and Entry. All scored data was counted and entered for each scoring procedure into an Excel spreadsheet by the author and a second individual independently counted and entered the data into a second spreadsheet. The second data-entry person was trained in CBM-W scoring, counting, and data. Once data were double counted and entered, the two spreadsheets were compared and any discrepancies were discussed and remediated on an individual basis. This process accounted for any counting and/or data-entry errors.

SAEBRS data was downloaded directly into Excel from Qualtrics. Following published scoring procedures, specific items were reverse coded using the formula function in Excel and 10% of all items were checked by hand to ensure reliability of reverse coding. Once all data had been double entered and/or reverse coded, where appropriate, all spreadsheets were merged into one data file using SPSS. Again, 10% of data was individually referenced with the original Excel spreadsheets to ensure data were merged correctly by study ID number.

Data Analysis

Reliability. Inter-rater reliability was calculated for all scoring procedures across both forms of CBM-W in order to determine scoring reliabilities. Pearson's correlations were used between forms A and B of each type of CBM-W within each grade for each time-point to determine alternate-form reliability.

Validity. Descriptive statistics for all CBM-W measures were reported, including the mean, standard deviation, range, skew, and kurtosis. Prior to analyzing concurrent and predictive validity, the ACCESS-W and MAP-ELA descriptive statistics were examined within each grade. Predictive and concurrent validity were explored using Pearson's and Spearman's correlations within each grade between CBM-W metrics and the ACCESS-W and for 3rd grade with the MAP-ELA. Additionally, correlations between CBM-W metrics were analyzed for divergent validity by comparing the correlations to the ACCESS-Literacy Composite (ACCESS-LC) to the ACCESS-Oral Language Composite (ACCESS-ORC) scores with the hypothesis being that CBM-W metrics should correlate more strongly with ACCESS-LC than ACCESS-ORC. For the MAP, correlations between CBM-W metrics, MAP ELA, and MAP-Math (MAP-MA) were

analyzed for divergent validity with the hypothesis being that CBM-W should correlated more strongly with MAP-ELA than MAP-MA.

Sensitivity to Growth. Means and standard deviations were calculated at every time-point by grade for each metric for WD and PW. A two-way mixed Analysis of Variance (ANOVA) test was used in conjunction with follow-up *t*-tests and repeated measures ANOVA (RM ANOVA) within grade to identify any significant differences between time and grade using the metrics with acceptable reliability ($r > .7$) and validity ($r > .4$). The between-subjects factor was grade and the within-subjects factor was time-point (fall, winter, spring).

EL Performance to General Population. In order to explore the differences between non-ELs and ELs, descriptive statistics from this study were compared to those from prior studies sampling from the general population using the same CBM-Ws from this study.

Predictive Validity/Regression. Hierarchical multiple regression was used with the metric showing the most promise for each CBM-W and the AB sub-scale of the SAEBRS. Specifically, all predictive hierarchical regressions were 3-model regressions with model-1 including the control variable of grade, model-2 adding the selected fall CBM-W metric, and model-3 adding the fall AB sub-scale score with the dependent variable being performance on the ACCESS-W. All concurrent hierarchical regressions were 4-model regressions with model-1 including the control variable of grade, model-2 adding Oral English Proficiency, model-3 adding selected winter CBM-W metrics, and model-4 adding the SAEBRS-AB for winter. Similar analyses were run for the MAP-ELA except they were only run for fall CBM-W and SAEBRS-AB to spring MAP-ELA

and winter CBM-W and SAEBRS-AB. Since only 3rd graders took the MAP-ELA, there was no need to control for grade as a step in the models. Therefore, fall was a 2-model regression with model-1 using fall CBM-W and model-2 adding fall SAEBRS-AB. Winter was 3-models with model-1 being Oral English Proficiency, then adding winter CBM-W, and finally SAEBRS-AB in winter. The primary purpose of the winter analyses was to examine the significance of including Oral English Proficiency in predictions. Prior to analyses, Cronbach's alpha was used to check the reliability of the various sub-scales of the SAEBRS.

Diagnostic Accuracy. A major purpose of universal screening using CBM is to identify students whom later fail or demonstrate problems. This predictive utility, or diagnostic accuracy, has been explored using state tests and other standardized assessments (Gersten et al., 2012; Jenkins, Hudson, Johnson, 2007). The Center on Response to Intervention (2014) noted that accuracy in classifying students as at-risk or not is central to technical adequacy. Effective screeners correctly discriminate between students who later have difficulties (true positives (TP)) and those with satisfactory performance (true negatives (TN)). Ineffective screeners produce false positives (FP; the screener indicates the student will fail but they do not) and false negatives (FN; the screener indicates the student will progress satisfactorily but they do not).

A screener's performance is often summarized according to *sensitivity* (correctly identifies TPs), *specificity* (correctly identifies TNs), and overall *classification accuracy* (TNs + TPs as identified by the screener/ Total Sample; i.e., percentage of total sample correctly identified as at-risk or no-risk). A minimal sensitivity level of .90 has been recommended for screeners because it is central to screening, early identification of those

in need of intervention in order to prevent failure (Jenkins et al., 2007), while minimal levels of specificity have been deemed adequate at .50 or better (Catts, Petscher, Schatschneider, Bridges, & Mendoza, 2009), .70 or better (Keller-Margulis et al., 2016), or .80 or higher (Compton et al., 2010). The result of lower sensitivity is FNs, meaning that students in need of early intervention are not identified early, while the result of lower specificity is FPs, meaning that students are provided early intervention when they did not necessarily need it to progress satisfactorily. When weighing the outcomes of lower sensitivity versus specificity, it seems to reason that high sensitivity (.90) satisfies the mission of MTSS and that specificity tolerance should be determined at a school level according to available resources (Clemens et al., 2011). However, Keller-Margulis and colleagues (2016) recommend evaluating CBM and the various metrics using a .70 threshold across both sensitivity and specificity. Due to disagreement surrounding ideal sensitivity and specificity, this study will examine cut-scores for each measure with sensitivity set as close to .90, .80, and .70 without going below and evaluate cut-score performance according to their respective specificity. Ideally, cut-scores will be found meeting .90 sensitivity and .70 specificity.

Logistic regression and ROC curves analyses have been used in CBM research to determine cut-scores using a measure of classification accuracy called the AUC (Clemens et al., 2011; Conoyer, Foegen, & Lembke, 2016; Johnson et al., 2009; Gersten et al., 2012). AUC values ranges from .5 (chance) to 1 (perfect). According to Clemens et al. (2011), AUC values of .90 or better are excellent, .80 to .89 are good, .70 to .79 are fair, and below .70 are poor. The Center of Response to Intervention (n.d.) considers an AUC of .85 or better to be “convincing evidence”. Thus, an AUC of .85 or better will be used

to provide an overall evaluation the various metrics. Logistic regression is used to obtain predicted probability values of combined screening measures, in this case either WD or PW and the SAEBRS-AB. According to Clemens and colleagues (2011), "predicted probabilities are an optimally weighted average of test scores that went into the logistic regression equation in regard to predicting group membership" (p. 236). Predicted probabilities represent an efficient way of combining measures to predict risk that has been used in several studies (Catts, Fey, Zhang, & Tomblin, 2001; Johnson, et al., 2010; and Clemens et al., 2011). The outcome variables of at-risk and no-risk were determined using the ACCESS-W and the MAP-ELA. Specifically, a scaled score of 2.9 or below was considered at-risk and 3.0 or above to be no risk for the ACCESS-W and a *Below Basic* was considered at-risk for the MAP-ELA.

Each participant was first coded as 1 ('at-risk) or 0 ('no-risk) according to criterion measure performance. Each predictor and the various combinations (WD-CLS and SAEBRS-AB; PW-WSC and SAEBRS-AB) were then entered into logistic regression equations to obtain predicted probabilities as a first step in determining whether or not SAEBRS-AB added significantly to WD-CLS and PW-WSC, respectively, in the prediction of risk as well as to obtain predicted probabilities for the combined predictors. ROC Curve analyses was then conducted with each predictor variable individually and each combined predictor using the predicted probabilities obtained via logistic regression. AUC was then used to provide an overall evaluation of each predictor but predictor performance approximating sensitivity levels of .90, .80, and .70 was also tabled to provide a more detailed evaluation of specific cut-score performance.

CHAPTER 4

RESULTS

Descriptive Statistics

Criterion Measures. *ACCESS.* The four sub-tests of the ACCESS (writing, reading, literacy composite, oral language composite) were examined for assumptions of normality via skew and kurtosis. As can be seen in Table 3, skew and kurtosis was acceptable for all subtests of the ACCESS for 2nd and 3rd grade while all were significantly skewed for 1st grade ($|z| > 1.96$, $p < .05$), indicating a non-normally distributed sample in 1st grade. Follow-up analysis using histograms and box-plots indicated the presence of an extreme outlier for the writing sub-test in 1st grade (beyond the 3rd quartile) and moderate outliers for the reading sub-test for 1st grade, the literacy composite for 1st grade, and the oral language composite for 2nd and 3rd grade.

Table 3

Descriptive Statistics for ACCESS

	Mean (SD)	Min	Max	Skew (Std. Er.)	Kurtosis (Std. Er.)
1st grade					
Writing	267(33.8)	214	371	1.35(.48)*	3.04(.94)*
Reading	288(38.6)	217	409	1.05(.47)*	3.23(.92)*
Literacy	277(32.7)	226	366	.94(.48)*	1.53(.94)
Oral Lang	318(50.2)	199	382	-.93(.47)*	.38(.92)
2nd grade					
Writing	293(27.3)	239	357	.09(.46)	-.01(.90)
Reading	313(32.3)	259	372	.47(.46)	-.68(.90)
Literacy	304(25.9)	265	365	.53(.46)	-.41(.90)
Oral Lang	293(53.6)	158	382	-.96(.46)	.95(.90)
3rd grade					
Writing	308(26.9)	262	362	.17(.48)	-.80(.94)
Reading	315(32.5)	264	370	.26(.47)	-1.13(.92)
Literacy	312(28.6)	267	363	.19(.48)	-1.25(.94)
Oral Lang	301(56.7)	176	373	-.97(.47)	-.03(.92)

* = $p < .05$

Shapiro-Wilk indicated a non-normal distribution ($p < .05$) for the ACCESS-W for 1st grade and the oral language composite (ACCESS-OLC) for 3rd grade. Further analysis of the extreme outlier indicated that student 62 was a 1st grade student with an overall composite score on the ACCESS as advanced. Participant 62 had the highest score across all grades on the ACCESS-W. Therefore, student 62 was removed and descriptive statistics were run again for 1st grade. As can be seen in Table 4, ACCESS-W was normally distributed according to skew and kurtosis once the outlier was removed. Shapiro-Wilk also indicated that the ACCESS-W was normal ($p = .76$) once the outlier was removed. The decision was made to remove participant 62 from further analysis.

Table 4

ACCESS 1st Grade Descriptive Statistics with Outlier Removed

	Mean (SD)	Min	Max	Skew (Std. Er.)	Kurtosis (Std. Er.)
1st grade					
Writing	262 (25.7)	214	323	.38 (.49)	.27 (.95)
Reading	287 (39.3)	217	409	1.12 (.48)*	3.22 (.94)*
Literacy	275 (30.5)	226	366	1.09 (.49)*	2.86 (.95)*
Oral Lang	316 (50.3)	199	382	-.88 (.48)	.32 (.94)

* = $p < .05$

After the outlier was removed, inter-correlation coefficients between the various sub-tests, including the ACCESS composite scores (ACCESS-Comp), within grade were computed in order to examine the relations between ACCESS-W and the various subtests. The purpose of examining inter-correlation coefficients was to determine if the ACCESS-W correlated more strongly with the ACCESS-LC than the ACCESS-OLC for purposes of divergent validity analysis for CBM-W. In other words, are ACCESS-LC and ACCESS-OLC measuring two highly similar or two different constructs. Results are

available in Table 5. ACCESS-W did correlate more strongly with ACCESS-LC than the ACCESS-OLC across grades.

Table 5

Correlations between ACCESS Sub-Tests

Variable	*ACCESS-R*	*ACCESS-LC*	*ACCESS-OLC*	*ACCESS-Comp*
ACCESS-W				
1st Grade	*.73***	*.90***	*.73***	*.92***
2nd Grade	*.51***	*.85***	*.50**	*.81***
3rd Grade	*.81***	*.94***	*.47**	*.84***

Note: * = $p < .05$; ** = $p < .01$, ACCESS-W = writing sub-test of ACCESS, ACCESS-R = reading subtest of ACCESS, ACCESS-LC = literacy composite score of ACCESS, ACCESS-OLC = oral language composite of ACCESS, ACCESS-Comp = composite ACCESS score

MAP-ELA. Descriptive statistics and correlations were calculated for the MAP-ELA and MAP-Math. Only 3rd grade participants completed the MAP in the spring, Table 6 provides descriptive statistics of student scores and correlations. Both MAP-ELA and MAP-Math appeared normally distributed according to skew, kurtosis, and examination of histograms. The two sub-tests were significantly correlated ($p < .01$) but not beyond $r > .70$. There were no extreme outliers.

Table 6

Descriptive Statistics and Inter-Correlations for MAP

	Mean (SD)	Min-Max	Skew	Kurtosis	Correlations
MAP-ELA N = 19	415.8 (41.1)	343-491	-.08 (.52)	-.83 (1.0)	.64**
MAP-Math N = 23	421.5 (42.5)	346-495	.03 (.48)	-.68 (.94)	.64**

Note: ** = $p < .01$, MAP-ELA = English Language Arts subtest of MAP

Predictor Measures. *Word Dictation (WD).* Inter-rater reliability (IRR) for scoring was 94% or better across all metrics for WD (WW, WSC, CLS, ILS) with a range of 94% - 99%. IRR was calculated for each metric at each time-point (fall, winter, spring). IRR for WW was 98%, 99%, and 99% respectively; 98%, 96%, and 98% for WSC respectively; CLS was 99%, 98%, and 99% respectively; and ILS was 96%, 94%, and 96% respectively. ILS had the lowest agreement percentages, likely because most students made fewer errors than CLS, which resulted in fewer scoring opportunities for ILS.

Descriptive statistics for mean performance on fall WD are available in Table 7. The *z* score of all skew and kurtosis statistics fell within +/- 1.96 except for WD-WSC for 1st grade ($z = 2.49$, $p < .05$). In general, a value within +/- 1.96 when dividing the skew or kurtosis by its respective standard error to convert it to a *z* score indicates a normal distribution (Rose, Spinks, & Canhoto, 2015). Furthermore, WD-WSC indicated possible floor effects for 1st graders (min = 0). Box plot analysis indicated moderate outliers in 1st grade for WSC and ILS, 2nd grade for WW and ILS, and 3rd grade for WW and ILS. However, none of the outliers were beyond the 3rd quartile.

Table 7

Descriptive Statistics for Fall WD

	Mean (SD)	Min	Max	Skew (Std. Er.)	Kurtosis (Std. Er.)
1st grade					
WW	17(6.5)	5	29	.09(.49)	-.95(.95)
WSC	5(4.5)	0	18	1.22(.49)*	1.41(.95)
CLS	49(27.6)	14	111	.60(.49)	-.52(.95)
ILS	24(11.0)	5	50	.66(.49)	.52(.95)
2nd grade					
WW	23(9.0)	2	44	-.26(.49)	1.04(.95)
WSC	12(10.1)	1	33	.69(.49)	-.69(.95)

CLS	86(49.4)	6	195	.37(.49)	-.47(.95)
ILS	25(15.6)	2	60	.43(.49)	-.21(.95)
3rd grade					
WW	26(7.7)	14	42	.30(.47)	-.06(.92)
WSC	15(17.9)	2	29	.27(.47)	-.94(.92)
CLS	103(39.6)	9	55	.31(.47)	-.39(.92)
ILS	27(10.2)	9	55	.73(.47)	.94(.92)

Note: * = $p < .05$

Descriptive statistics for winter administration of WD can be found in Table 8. WSC was significantly skewed for 1st grade ($z = 2.47$, $p < .05$) and ILS was significantly skewed for 1st grade ($z = 6.02$, $p < .001$), 2nd grade ($z = 2.70$, $p < .01$), and 3rd grade ($z = 2.21$, $p < .05$). ILS also had significant kurtosis for grades 1 and 2 ($p < .05$). WW and CLS appeared normally distributed across all three grades. Analysis of box plots indicated a presence of outliers for ILS at each grade, two for WSC at first grade, and outliers for WW in 2nd grade and 3rd grade. However, no outliers appeared extreme, beyond the 3rd quartile. Further analysis using histograms did not indicate extreme outliers for WW, WSC, or CLS but did show evidence of extreme outliers for 1st and 2nd grade ILS. Additional tests of normality were run using Shapiro-Wilk. WW, CLS, and ILS were not significant ($p > .05$) across grades while WSC was not significant for 3rd grade but was significant for 1st grade ($p = .003$) and second grade ($p = .041$). Thus, Shapiro-Wilk test indicated that WSC was not normally distributed for 1st and 2nd grade, so any further analysis should be interpreted with caution.

Table 8

Descriptive Statistics for Winter WD

	Mean (SD)	Min	Max	Skew (Std. Er.)	Kurtosis (Std. Er.)
1st grade					
WW	20(6.7)	10	32	.50 (.50)	-.69 (.97)
WSC	6(4.9)	1	18	1.24 (.50)*	.54(.97)

CLS	57 (29.2)	15	120	.68 (.50)	-.36 (.97)
ILS	31 (22.6)	11	116	2.93(.50)***	10.13(.97)***
2nd grade					
WW	27(9.3)	10	47	.18(.47)	-.47(.92)
WSC	15(10.5)	3	35	.37(.47)	-1.13(.92)
CLS	105(51.8)	31	211	.34(.47)	-.94(.92)
ILS	28(21.6)	3	92	1.27(.47)**	1.77(.92)*
3rd grade					
WW	31(7.4)	17	45	-.04(.47)	-.15(.92)
WSC	17(8.4)	3	34	.12(.47)	-.93(.92)
CLS	121(40.4)	38	203	.07(.47)	-.35(.92)
ILS	35(17.1)	13	75	1.04(.47)*	.44(.92)

*Note: * = $p > .05$; ** $p > .01$; *** = $p > .001$*

Descriptive statistics for spring administration of WD can be found in Table 9. ILS was significantly skewed for 1st grade (z = 6.90, $p < .01$), 2nd grade ($z = 3.26$, $p < .01$), and 3rd grade ($z = 2.63$, $p < .01$). ILS also had significant kurtosis at 1st grade ($z = 21.11$, $p < .01$) and 3rd grade ($z = 3.16$, $p < .01$). However, all other metrics appeared normally distributed across grades. Analysis of box plots indicated the presence of an extreme outlier for ILS in 1st grade, two moderate outliers for ILS in 2nd grade, and one moderate outlier for ILS in 3rd grade. There were not outliers for any of the other metrics. Further analysis using histograms confirmed the results of the box plots. Additional tests of normality were run using Shapiro-Wilk. For 1st grade, Shapiro-Wilk indicated that WSC was not normally distributed ($p = .017$) as well as ILS ($p < .001$). For 2nd grade, Shapiro-Wilk indicated that WSC ($p = .042$), CLS ($p = .032$), and ILS ($p = .002$) were not normally distributed. For 3rd grade, Shapiro-Wilk indicated that all metrics were normally distributed. Thus, Shapiro-Wilk test indicated that only WW was normally distributed across grades, so any further analysis should be interpreted with caution.

Table 9

Descriptive Statistics for Spring WD

	Mean (SD)	Min	Max	Skew (Std. Er.)	Kurtosis (Std. Er.)
1st grade					
WW	22.6 (8.1)	9	42.5	.67 (.51)	.45 (.99)
WSC	7.7 (6.5)	0	21.5	.95 (.51)	-.26 (.99)
CLS	67.2 (36.1)	7	146	.61 (.51)	-.12 (.99)
ILS	34.6 (33.4)	8.5	166.5	3.52 (.51)**	14.07 (.99)**
2nd grade					
WW	30.7 (11.3)	14.5	56	.59 (.46)	-.43 (.90)
WSC	19.1 (12.6)	3.5	43.5	.58 (.46)	-.61 (.90)
CLS	122.5 (62.7)	46	247.5	.64 (.46)	-.70 (.90)
ILS	28.8 (25.1)	2	100.5	1.50 (.46)**	1.77(.92)
3rd grade					
WW	32.8 (9.3)	14.5	56	.48 (.49)	-.03 (.95)
WSC	20.7 (9.8)	5.5	43.5	.54 (.49)	-.01 (.95)
CLS	138.1 (49.2)	64.5	251	.55 (.49)	.02 (.95)
ILS	28.6 (13.5)	7	70	1.29 (.49)**	3.00 (.95)**

*Note: * = p > .05; ** p > .01*

Picture Word (PW). The percentage of IRR for scoring PW-WW was 99% across all time-points, 98% for WSC, 96%-98% for CWS, and 88%-91% for IWS. Thus, IWS and metrics incorporating IWS (i.e., C-IWS) are more prone to issues related to scoring reliability. Descriptive statistics for mean performance on fall PW are available in Table 10. The *z*-scores of all skew and kurtosis statistics fell within +/- 1.96 except for second grade. For 2nd grade, skew and kurtosis, respectively, were WW ($z = 3.50$, $p < .01$; $z = 2.93$, $p < .01$), WSC ($z = 3.38$, $p < .01$; $z = 2.79$, $p < .01$), and CWS ($z = 2.27$, $p < .05$; $z = 1.49$, $p > .05$). Box plot analysis indicated moderate outliers in 2nd grade across metrics (WW, WSC, CWS) as well as one extreme outlier for WSC. Follow-up analyses using histograms also indicated a possible outlier for WSC in 2nd grade. However, it was decided to keep the outlier because of the small sample size and CWS was not significantly skewed. Shapiro-Wilk indicated that WW ($p = .001$), WSC ($p = .001$), and CWS ($p = .030$) were not normally distributed for 2nd grade, but all other grades were $p > .05$.

Table 10

Descriptive Statistics for Fall PW

	Mean (SD)	Min	Max	Skew (Std. Er.)	Kurtosis (Std. Er.)
1st grade					
WW	15 (6.3)	3	23	-.42(.50)	-1.17(.97)
WSC	11(5.0)	3	19	.00(.50)	-.93(.97)
CWS	9(4.8)	1	18	.35(.50)	-.56(.97)
2nd grade					
WW	30(14.6)	13	70	1.68(.48)**	2.75(.94)**
WSC	27 (13.8)	12	66	1.62(.48)**	2.62(.94)**
CWS	24(10.8)	8	53	1.09(.48)*	1.40(.94)
3rd grade					
WW	28 (13.5)	8	57	.60(.48)	-.13(.94)
WSC	25(12.9)	7	55	.76(.48)	.08(.94)
CWS	23(13.6)	4	58	.81(.48)	.42(.94)

Note: * = $p < .05$; ** = $p < .01$

Descriptive statistics can be found in Table 11 for winter PW. WW was significantly skewed for 2nd grade ($z = 2.91$, $p < .01$) and WSC was significantly skewed for 2nd grade ($z = 2.19$, $p < .05$). CWS appeared normally distributed across all three grades. Analysis of box plots indicated a presence of outliers for WW and CWS for 2nd grade. However, no outliers appeared extreme, beyond the 3rd quartile. One outlier was present for WW in 3rd grade but was not extreme. Further analysis using histograms did not indicate extreme outliers for WW, WSC, or CWS. WW and WSC were not normally distributed according to Shapiro-Wilk ($p < .05$) for 2nd grade, but all other metrics were normally distributed across grades ($p > .05$).

Table 11

Descriptive Statistics for Winter PW

	Mean (SD)	**Min**	**Max**	**Skew (Std. Er.)**	**Kurtosis (Std. Er.)**
1st grade					
WW	17 (6.6)	4	31	-.19(.48)	-.22(.94)
WSC	14(5.1)	4	24	.05(.48)	-.06(.94)

CWS	11(5.8)	1.5	24	.07(.48)	-.44(.94)
2nd grade					
WW	32(14.3)	16	72	1.34(.46)**	1.73(.90)
WSC	29(12.8)	12.5	61.5	1.01(.46)*	.65(.90)
CWS	27(10.0)	10.5	49	.38(.46)	-.27(.90)
3rd grade					
WW	32(10.2)	13.5	58	.89(.48)	.91(.94)
WSC	30(9.8)	13.5	52.5	.84(.48)	.35(.94)
CWS	29(11.3)	9	56	.57(.48)	.16(.94)

*Note: *= p > .05; ** p > .01*

Descriptive statistics can be found in Table 12 for spring PW. None of the metrics had significant skew or kurtosis across grades. Analysis of box plots indicated a presence of one moderate outlier for CWS in 1st grade. Further analysis using histograms did not indicate extreme outliers. Shapiro-Wilk indicated a normal distribution across metrics for all grades ($p > .05$). Thus, spring PW metrics were normally distributed.

Table 12

Descriptive Statistics for Winter PW

	Mean (SD)	**Min**	**Max**	**Skew (Std. Er.)**	**Kurtosis (Std. Er.)**
1st grade					
WW	22 (1.7)	8	34.5	-.05 (.51)	-.84 (.99)
WSC	18.2 (1.6)	6.5	32	.40 (.51)	-.40(.99)
CWS	16.7 (2.2)	.5	37.5	.61 (.51)	.16 (.99)
2nd grade					
WW	36.5 (3.0)	12.5	75	.48 (.46)	.19 (.90)
WSC	34.3 (2.8)	10.5	60.5	.04 (.46)	-.97 (.90)
CWS	35.2 (3.1)	8.5	61	.09 (.46)	1.02 (.90)
3rd grade					
WW	38.0 (2.8)	10.5	56	-.36 (.49)	-1.02 (.95)
WSC	35.3 (2.8)	9.5	53	-.34 (.49)	-1.22 (.95)
CWS	35.8 (3.4)	10.5	62	-.12 (.49)	-1.29 (.95)

*Note: *= p > .05; ** p > .01*

Social, Academic, and Emotional Behavior Risk Screener (SAEBRS). Descriptive statistics for the various sub-scales and composite score of the SAEBRS are provided in Table 13. Examination of mean scores across time-points for each sub-scale within grade

reveals relative consistency of scores across times as well as across grades, with the exception of 3rd grade students, which consistently had a mean score below those of 1st and 2nd grade students in the SAEBRS-AB sub-scale and total score.

Table 13

Means and Standard Deviations of the SAEBRS

	AB			Social			Emo			Total		
	Fall	Win	Sp	Fall	Win	Sp	Fall	Win	Sp	Fall	Win	Sp
1st	13	15	15	15	16	16	17	18	18	45	49	49
	(3.5)	(3.2)	(3.7)	(3.0)	(2.6)	(3.0)	(2.9)	(3.0)	(3.2)	(6.9)	(7.9)	(9.0)
2nd	13	14	15	16	17	17	18	19	17	46	49	49
	(3.5)	(3.6)	(2.6)	(2.5)	(1.9)	(2.0)	(2.9)	(2.5)	(3.6)	(7.3)	(6.1)	(6.5)
3rd	12	13	12	15	15	15	17	18	17	44	46	44
	(3.7)	(3.8)	(3.9)	(3.3)	(3.6)	(4.4)	(3.0)	(3.4)	(3.7)	(8.4)	(9.5)	(10.7)

Note: AB = academic behavior subscale of SEABRS, Emo = Emotional behavior subscale of SAEBRS

The published *at-risk* cut score for SAEBRS-AB is 9 or below, 12 or below for Social Behavior, 17 or below for Emotional Behavior, and 36 or below for Total Behavior. Table 14 provides the number and percentage of students scoring as *at-risk* within each subscale at each time-point within each grade. Percentages of *at-risk* are higher in fall for 1st and 2nd grade than in spring but *at-risk* rates increased from fall to spring for 3rd grade in SAEBRS-AB. Rates of risk in SAEBRS-AB for 2nd grade fell from 20% in fall to 0% in spring. Across all grades and time-points, large percentages of student were *at-risk* in Emotional Behavior. In general, 3rd grade participants had higher rates of risk than 1st or 2nd grade participants.

Table 14

Rates of Risk According to SAEBRS

	AB			Social			Emo			Total		
	Fall	Win	Sp	Fall	Win	Sp	Fall	Win	Sp	Fall	Win	Sp
1st	3	1	1	4	3	3	13	9	5	1	2	2
	13%	4%	4%	17%	13%	13%	57%	39%	22%	4%	9%	9%

2nd	5	4	0	3	2	2	11	10	11	3	1	1
	20%	16%	0%	12%	8%	8%	44%	40%	44%	12%	4%	4%
3rd	5	5	6	5	6	4	14	11	10	4	6	6
	21%	21%	25%	21%	25%	17%	58%	46%	42%	17%	25%	25%

Note: AB = academic behavior subscale of SEABRS, Emo = Emotional behavior subscale of SAEBRS

Internal reliability for the various sub-scales was assessed using Cronbach's alpha (α). For reference, internal consistency of the SAEBRS sub-scales reported with the general population of elementary school students is .90-.92 for SAEBRS-AB, .89-.94 for Social Behavior, .83 for Emotional Behavior, and .93 for Total Score. Table 15 provides α for each sub-scale within each grade and for each time-point for the participants in this study. For the participants in this study, α ranged from .85-.95 for SAEBRS-AB, .74-.93 for Social Behavior, and .72-.91 for Emotional Behavior. The lowest α across subscales was consistently in the fall, especially for 2nd graders. However, the scale of most importance to this study is the SAEBRS-AB sub-scale and internal consistency was acceptable ($\alpha > .80$) across grades and time-points.

Table 15

Internal-Consistency of SAEBRS Sub-Scales

	AB			**Social**			**Emo**		
	Fall	Win	Sp	Fall	Win	Sp	Fall	Win	Sp
1st	.93	.92	.95	.81	.88	.89	.81	.90	.91
2nd	.85	.91	.89	.74	.87	.84	.73	.83	.91
3rd	.90	.89	.91	.87	.89	.93	.72	.90	.90

Note: AB = academic behavior subscale of SEABRS, Emo = Emotional behavior subscale of SAEBRS

Summary. IRR was strong for all metrics except for those including measures of incorrect sequences. One 1st grade participant was removed from analyses due to extremely high performance on the ACCESS-W, but both criterion measures (ACCESS-W and MAP-ELA) were normally distributed once the outlier was removed. There were concerns with normal distributions (skew) for incorrect sequences (WD-ILS and PW-IWS) as well as for IRR. There were also concerns related to normal distribution (skew) and possible floor effects for WD-WSC and PW-WSC. Concerns related to normal distributions and small sample size warrants parallel parametric and non-parametric analyses for validity. The purpose of running both parametric and non-parametric analyses was to determine if results were relatively equitable across analyses. Extreme differences between parametric and non-parametric analyses indicate concerns related to the validity of findings.

Reliability

Word Dictation. Alternate-form reliability for each time point, grade, and scoring procedure is provided in Table 16. According to McMaster and Espin (2007), reliability coefficients of $r > .80$ are strong, $r = .70$ to .79 are moderately strong and sufficient for further analysis, and $r < .70$ are unacceptable. WD-ILS was the only scoring procedure not meeting the alternate-form threshold of .70 for 2nd and 3rd graders in the fall and 3rd graders in the spring. WD-CLS met $r = .91$ or higher for each grade at each time-point. WD-WSC also met $r = .91$ or better for all time-points and grades except for 3rd grade in winter ($r = .88$). WD-WW also performed well, especially in winter and spring ($r = .81$ - .97). Thus, WD-WW, WSC, and CLS were worthy of further examination, but WD-ILS was removed from further analyses.

Table 16

Alternate Form Reliability for Word Dictation

	Fall			**Winter**			**Spring**		
	1st	**2nd**	**3rd**	**1st**	**2nd**	**3rd**	**1st**	**2nd**	**3rd**
WW	.81**	.87**	.87**	.94**	.92**	.88**	.97**	.96**	.97**
WSC	.97**	.91**	.91**	.92**	.93**	.88**	.92**	.97**	.92**
CLS	.95**	.93**	.91**	.95**	.94**	.91**	.97**	.97**	.94**
ILS	.73**	.63**	.65**	.94**	.88**	.84**	.98**	.92**	.61**

** $p < .001$

Picture Word. Alternate form reliability data for PW are available in Table 17 and was calculated and evaluated using the same criteria as for WD-CBM. Alternate form reliability ranged from weak to strong (r = .60 to .97). PW-WW and WSC performed above the .70 criteria across grades and time-points, PW-CWS met the criteria for all grades and time-points except for 1st grade in the fall (r = .60). Therefore, PW-WW and/or WSC appeared to produce more reliable data in the fall for 1st grade.

Table 17

PW Alternate Form Reliability

	Fall			Winter			Spring		
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
WW	.81**	.90**	.91**	.88**	.94**	.81**	.77**	.91**	.97**
WSC	.72**	.89**	.92**	.85**	.92**	.83**	.71**	.88**	.96**
CWS	.60**	.72**	.88**	.91**	.82**	.84**	.86**	.88**	.94**

** = $p < .01$; * = $p < .05$

Criterion Validity

ACCESS-W. Both concurrent and predictive criterion validity were examined for WD, PW, and SAEBRS-AB with the ACCESS-W via parallel parametric (Pearson's) and non-parametric (Spearman's Rho) correlations. Concurrent validity was examined by correlating winter predictors to the ACCESS-W, which was administered in January, and predictive validity was examined via correlating fall predictors to the ACCESS-W. The guidelines used to evaluate the quality of correlations were drawn from standards set by Marston (1985) for CBM research ($r \geq .70$ = strong; $r = .50$ to .69 = moderate; and $r \leq .50$ = weak).

Concurrent Validity. First, scatterplots were examined between the winter WD-WW, WSC, CLS, and the ACCESS-W for each grade for evidence of a linear relationship. Visual analysis of first grade revealed clustering at the low end of ACCESS-W performance and WD-WSC did not have a clear linear relation with the ACCESS-W for first grade. WD-WW and CLS appeared to have a linear relation with the ACCESS-W for first graders. Linear relations were evident for WD-WW, WSC, and CLS for 2nd and 3rd grade. Next, scatterplots were examined between winter PW-WW, WSC, CWS, and the ACCESS-W for each grade for evidence of a linear relationship. Visual analysis indicated the presence of a linear relationship across grades for PW-WW, WSC, and CWS. Finally, scatterplots of the winter SAEBRS-AB were examined with the ACCESS-W. Visual analysis indicated a linear relationship for 2nd and 3rd grade but clustering for 1st grade. Pearson and Spearman's Rho correlations by grade are available in Tables 18-20.

Table 18

1st Grade Concurrent Validity Correlations to ACCESS-W

Variable	1	2	3	4	5	6	7	8
1. ACCESS-W	-	*.30*	*.64***	*.50**	*.23*	*.22*	*.05*	*.18*
2. WD-WW	.26	-	*.48**	*.69***	*.85***	*.55***	*.45**	*-.03*
3. WD-WSC	.57**	.48*	-	*.90***	*.37*	*.41*	*.40*	*.42*
4. WD-CLS	.56*	.64**	.92**	-	*.60***	*.65***	*.58***	*.33*
5. PW-WW	.19	.84**	.42	.56**	-	*.79***	*.62***	*.08*
6. PW-WSC	.19	.57**	.46*	.58**	.87**	-	*.89***	*.32*
7. PW-CWS	.15	.39	.48*	.52*	.67**	.88**	-	*.50**
8. SAEBRS-AB	.30	.07	.35	.37	.16	.31	.51*	-

Note: Correlations below the diagonal are Pearson's and those above are Spearman's Rho; * = $p < .05$; ** = $p < .01$

Table 19

2nd Grade Concurrent Validity Correlations to ACCESS-W

Variable	1	2	3	4	5	6	7	8
1. ACCESS-W	-	*.68***	*.82***	*.78***	*.50**	*.52**	*.68***	*.27*
2. WD-WW	.62**	-	*.72***	*.87***	*.79***	*.78***	*.74***	*-.06*
3. WD-WSC	.81**	.71**	-	*.92***	*.42**	*.44**	*.57***	*.34*
4. WD-CLS	.78**	.87**	.95**	-	*.58***	*.58***	*.67***	*.07*
5. PW-WW	.57**	.80**	.53**	.64**	-	*.99***	*.89***	*-21*
6. PW-WSC	.57**	.80**	.53**	.63**	.99**	-	*.90***	*-.14*
7. PW-CWS	.62**	.73**	.56**	.62**	.84**	.88**	-	*-.01*
8. SAEBRS-AB	.42*	-.02	.35	.21	.00	.04	.21	-

Note: Correlations below the diagonal are Pearson's and those above are Spearman's Rho; * = $p < .05$; ** = $p < .01$

Table 20

3rd Grade Concurrent Validity Correlations to ACCESS-W

Variable	1	2	3	4	5	6	7	8
1. ACCESS-W	-	*.56***	*.67***	*.64***	*.42**	*.42**	*.40*	*.36*
2. WD-WW	.63**	-	.67**	.84**	.50*	.49*	.53**	.56**
3. WD-WSC	.69**	.70**	-	*.94***	*.58***	*.59***	*.62***	*.40*
4. WD-CLS	.68**	.89**	.94**	-	*.54***	*.54***	*.58***	*.44**
5. PW-WW	.48*	.58**	.62**	.64**	-	*.99***	*.94***	*.32*
6. PW-WSC	.50*	.59**	.65**	.66**	.99**	-	*.95***	*.34*
7. PW-CWS	.48*	.61**	.67**	.68**	.95**	.97**	-	*.27*
8. SAEBRS-AB	.43*	.52**	.44*	.50*	.34	.38	.31	-

Note: Correlations below the diagonal are Pearson's and those above are Spearman's Rho; * = $p < .05$; ** = $p < .01$

For 1st grade, WD metrics had weak to moderate (.26-.57) correlations, PW had weak correlations (.15-.19), and the SAEBRS-AB correlated weakly (.30) with ACCESS-W. Correlations remained consistent across both parametric and non-parametric analyses. WD-WSC and CLS correlated moderately with ACCESS-W (r = .57 and .56, respectively) and had the best concurrent validity for 1st grade. For 2nd grade, WD metrics had moderate to strong correlations (r = .62-.81), PW metrics were moderately correlated (r = .57-.62), and SAEBRS-AB correlated weakly with the ACCESS-W (r = .42). Again, correlations were consistent across parametric and non-parametric analyses. As was the case for 1st grade, WD-WSC and CLS appeared to have the strongest concurrent validity with the ACCESS-W (r = .81 and .78, respectively). For 3rd grade, WD metrics had strong correlations (r = .63-.69), PW metrics had weak correlations (r = .48-.50), and SAEBRS-AB correlated weakly with ACCESS-W (r = .43). Correlations were consistent across parametric and non-parametric analyses. WD-WSC and CLS had the strongest correlations with the ACCESS-W for 3rd grade (r = .69 and .68, respectively) and therefore across all grades. PW-WW, WSC, and CWS correlated relatively equitably within each grade to ACCESS-W, so there was no clearly superior metric for PW.

Predictive Validity. Visual analysis of scatterplots of fall WD-WW with ACCESS-W revealed linear relationships across grades, as did WD-CLS. WD-WSC had a clear linear relationship in the 2nd and 3rd grade but there was clustering at the low end for 1st grade. Visual analysis of scatterplots of winter predictors to ACCESS-W indicated the presence of a linear relationship across grades for PW-WW, WSC, CWS, and SAEBRS-AB. As with concurrent validity, both Pearson and Spearman's Rho

correlations were run for predictive validity. Pearson and Spearman's Rho correlations by grade are available in Tables 21-23.

Table 21

1st Grade Predictive Validity Correlations for Fall Predictors to ACCESS-W

Variable	1	2	3	4	5	6	7	8
1. ACCESS-W	-	*.40*	*.65***	*.54**	*.43*	*.31*	*.27*	*.48**
2. WD-WW	.43	-	*.75***	*.92***	*.90***	*.74***	*.59***	*.33*
3. WD-WSC	.65**	.75*	-	*.92***	*.69***	*.61***	*.54**	*.58***
4. WD-CLS	.61**	.90**	.95**	-	*.84***	*.73***	*.62***	*.47**
5. PW-WW	.43	.86**	.64**	.78**	-	*.94***	*.82***	*.30*
6. PW-WSC	.42	.74**	.61**	.70**	.92**	-	*.91***	*.23*
7. PW-CWS	.38	.62**	.57**	.62**	.76**	.92**	-	.22
8. SAEBRS-AB	.42	.31	.38	.42	.34	.23	.16	-

*Note***:** Correlations below the diagonal are Pearson's and those above are Spearman's Rho; * = $p < .05$; ** = $p < .01$

Table 22

2nd Grade Predictive Validity Correlations for Fall Predictors to ACCESS-W

Variable	1	2	3	4	5	6	7	8
1. ACCESS-W	-	*.56***	*.70***	*.67***	*.43**	*.47**	*.52**	*.36*
2. WD-WW	.63**	-	*.78***	*.86***	*.77***	*.78***	*.80***	*.38*
3. WD-WSC	.73**	.77**	-	*.98***	*.45**	*.45**	*.65***	*.49**
4. WD-CLS	.72**	.89**	.97**	-	*.55***	*.55***	*.71***	*.45**
5. PW-WW	.61**	.75**	.60**	.67**	-	*.99***	*.86***	*.34*
6. PW-WSC	.62**	.76**	.61**	.67**	.99**	-	*.86***	*.34*
7. PW-CWS	.65**	.80**	.76**	.80**	.91**	.92**	-	.38
8. SAEBRS-AB	.35	.37	.54*	.48*	.37	.37	.42*	-

Note: Correlations below the diagonal are Pearson's and those above are Spearman's Rho; * = $p < .05$; ** = $p < .01$

Table 23

3rd Grade Predictive Validity Correlations for Fall Predictors to ACCESS-W

Variable	1	2	3	4	5	6	7	8
1. ACCESS-W	-	*.59***	*.71***	*.66***	*.47**	*.51**	*.52**	*.42**
2. WD-WW	.64**	-	.79**	.90**	.62**	.63**	.67**	.26
3. WD-WSC	.69**	.86**	-	*.96***	*.58***	*.63***	*.70***	*.37*
4. WD-CLS	.68**	.96**	.96**	-	*.55***	*.60***	*.66***	*.30*
5. PW-WW	.53**	.70**	.66**	.70**	-	*.99***	*.93***	*.06*
6. PW-WSC	.56**	.72**	.70**	.73**	.99**	-	*.94***	*.07*
7. PW-CWS	.56**	.73**	.72**	.74**	.94**	.97**	-	*.12*

8. SAEBRS-AB	.42*	.38	.44*	.41*	.20	.22	.24	-

Note: Correlations below the diagonal are Pearson's and those above are Spearman's Rho; * = $p < .05$; ** = $p < .01$

For 1st grade, WD metrics had weak to moderate (.43-.65) correlations, PW had weak correlations (.38-.43), and the SAEBRS-AB correlated weakly (.42) with ACCESS-W. PW correlations were higher from fall to ACCESS-W than from winter to ACCESS-W. The low alternate-form reliability for fall PW-CWS in 1st grade coupled with discrepancies between Pearson's and Spearman's Rho for PW-CWS and WSC raise doubt regarding significance of these results. However, correlations remained consistent across both parametric and non-parametric analyses for the WD metrics, PW-WW, and the SAEBRS-AB.

WD-WSC and CLS correlated moderately with ACCESS-W ($r = .61$ and .65, respectively) and had the best predictive validity for 1st grade. For 2nd grade, WD metrics had moderate to strong correlations with ACCESS-W ($r = .63$-.73), PW metrics were moderately correlated with ACCESS-W ($r = .61$-.65), and SAEBRS-AB correlated weakly with the ACCESS-W ($r = .35$). Again, correlations were consistent across parametric and non-parametric analyses across all measures except PW-WSC and CWS. As was the case for 1st grade, WD-WSC and CLS appeared to have the strongest predictive validity with the ACCESS-W ($r = .73$ and .72, respectively). For 3rd grade, WD metrics had moderate correlations ($r = .64$-.69), PW metrics had moderate correlations ($r = .53$-.56), and SAEBRS-AB correlated weakly with ACCESS-W ($r = .42$). Correlations were consistent across parametric and non-parametric analyses, including PW-WSC and CWS. WD-WSC and CLS had the strongest correlations with the ACCESS-W for 3rd grade ($r = .69$ and .68, respectively) and therefore across all

grades. PW-WW, WSC, and CWS correlated relatively equitably within each grade to ACCESS-W, so there was no clearly superior metric for PW. These findings were consistent with concurrent validity.

Divergent Validity. Another way to examine validity is whether or not a predictor correlates more strongly with a criterion measure of the intended construct (writing, literacy) and not as strongly with a measure of a different construct (oral language, math). For the ACCESS, the two constructs to be compared are literacy and oral language. Although oral language and literacy are highly related, it is still expected that CBM-W would correlate more strongly with the ACCESS-LC than the ACCESS-ORC. Results for the various predictors to both ACCESS-LC and ACCESS-ORC are available in Table 24, using Pearson's correlations. WD-WSC, WD-CLS, PW-WSC, and PW-CWS consistently correlated more strongly with ACCESS-LC than ACCESS-ORC. WD-WW correlated more strongly with ACCESS-ORC in fall and winter for 1st grade while PW-WW correlated more strongly with ACCESS-ORC in winter for 1st grade and fall for 3rd grade. SAEBRS-AB correlated more strongly with ACCESS-ORC in fall for 1st grade and winter for 2nd grade.

Table 24

Divergent Validity of Predictors to ACCESS-LC and ORC

	1st Grade				2nd Grade				3rd Grade			
	Fall		Winter		Fall		Winter		Fall		Winter	
Predictor	LC	ORC	LC	ORC	LC	ORC	LC	ORC	LC	ORC	LC	ORC
WD-WW	.38	.41	.20	.32	.69**	.46*	.58**	.18	.66**	.55**	.66**	.13
WD-WSC	.67**	.55**	.62**	.48*	.78**	.57**	.82**	.46*	.68**	.56**	.67**	.44*
WD-CLS	.60**	.52*	.54*	.49*	.77**	.55**	.75**	.39	.68**	.58**	.68**	.33
PW-WW	.41	.30	.17	.21	.62**	.10	.57**	.06	.49*	.52*	.50*	.34
PW-WSC	.48*	.16	.23	.12	.65**	.11	.59**	.08	.52*	.48*	.51*	.29
PW-CWS	.44	.15	.20	.06	.73**	.29	.68**	.38	.53**	.50*	.49*	.29
SAEBRS-AB	.18	.32	.18	.10	.35	.34	.29	.59**	.52*	.22	.49*	-.06

Note: * = $p < .05$; ** = $p < .01$; LC = literacy composite of ACCESS; ORC = oral language composite of ACCESS

Summary. WD-WSC and CLS consistently had the highest concurrent and predictive correlations with the ACCESS-W across grades (r = .43-.81). However, concerns with possible floor effects and normal distribution for WD-WSC indicate WD-CLS may be the most robust measure as it relates to ACCESS-W. Of the PW metrics, no single metric stood out by consistently performing more strongly than the other. However, results from divergent validity indicate that metrics incorporating accuracy are more discriminative than production only metrics. Therefore, PW-WSC and CWS appear to be the most valid PW metrics as determined by the ACCESS. PW performed best for 2nd grade (r = .57-.65) and was acceptable for 3rd grade (r = .48-.56) but did not perform well for 1st grade (r = .15-.43). Beyond PW having weak correlations with the ACCESS-W for 1st grade, PW-CWS also failed to meet alternative form reliability acceptability guidelines in fall for 1st grade.

MAP-ELA. The second criterion measure used to examine CBM-W and SAEBRS-AB validity was the MAP-ELA, but only for 3rd grade. Spring administration of the predictors was used for concurrent validity while winter and fall administrations were used for predictive validity. Both parametric and non-parametric correlations are used to maintain consistency of analyses across criterion measures.

Concurrent Validity. First, scatterplots were used to determine whether or not linear relationships existed between the various predictors and the MAP-ELA for spring administration of predictors. Evidence of a linear relationship was present between each predictor in spring and the MAP-ELA scaled score. Results of the correlations are presented in Table 25. Note that only correlations between the predictor and the MAP-ELA are presented in Table 25 because correlations between the various predictors have

already been presented in Tables 18-23. For concurrent validity, WD metrics were weakly correlated with the MAP-ELA (r = .43-.45), PW metrics were moderately related to MAP-ELA (r = .66-.67), and SAEBRS-AB was moderately correlated with the MAP-ELA (r = .63). Relationships were similar across parametric and nonparametric analyses. For the MAP-ELA, PW metrics and the SAEBRS-AB were more strongly related to correlated than the WD metrics and there were negligible differences between the metrics for each respective CBM-W. In other words, production only metrics (WW) were similar to those including accuracy (WSC, CLS, CWS) for their respective CBM-W (WD and PW).

Predictive Validity. Evidence of a linear relationship was present between each predictor and the MAP-ELA scaled score. Results are presented in Table 25. For predictive validity, WD metrics were weak (r = .33 - .50), PW metrics were weak to moderate (r = .48-.56), and SAEBRS-AB was weak to moderate (r = .45-.62). For WD, there was little difference between the various metrics at each time-point but there were discrepancies between parametric and nonparametric analyses in fall for WD-WW (r = .33; r_s = *.19).* For PW, analyses were consistent across parametric and non-parametric correlations and the various metrics performed relatively equitably. Again, PW and SAEBRS-AB had the strongest correlations with the MAP-ELA (r = .44-.67).

Table 25
Correlations Between Predictors and MAP-ELA for 3rd Grade.

Predictor	Fall		Winter		Spring	
	r	r_s	r	r_s	r	r_s
WD-WW	.33	.19	.50*	.46*	.45	.51*
WD-WSC	.43	.42	.43	.41	.44	.47
WD-CLS	.35	.26	.44	.37	.43	.46
PW-WW	.48*	.43	.54*	.49*	.65**	.62**
PW-WSC	.50*	.44	.56*	.51*	.66**	.63**

PW-CWS	.44	.40	.46*	.44	.67**	.66**
SAEBRS-AB	.45	.50*	.62**	.58**	.63**	.67**

Note: * = $p < .05$; ** = $p < .01$; r = Pearson's Coefficient; r_s =Spearman's Rho

Divergent Validity. Analyses and comparisons were made with MAP just as they were with the ACCESS with the exception that the other construct assessed by MAP was mathematics via the MAP-MA. CBM-W should correlate more strongly with MAP-ELA than MAP-MA because math and writing/literacy are two separate constructs. Results are presented in Table 26. For fall and winter, there were large differences between correlations between all CBM-W metrics and the MAP-ELA (r = .33 to .56) and MAP-MA (r = -.03 to .22), indicating both forms of CBM-W were more strongly related to literacy than mathematics in the fall and winter. However, WD metrics had stronger correlations with the MAP-MA (r = .50 to .51) than the MAP-ELA (r = .43 to .45) in spring. PW metrics were more strongly correlated with MAP-MA in spring (r = .42 to .53) than in either fall or winter, but were still more strongly correlated to MAP-ELA (r = .65 to .67) than MAP-MA. SAEBRS-AB was more strongly correlated to MAP-ELA than MAP-MA in the fall and spring but not in the winter. SAEBRS-AB also correlated very similarly across criterion measures within each time-point. PW was the only predictor measure that correlated more strongly with MAP-ELA than the MAP-MA across time-points.

Table 26

Divergent Validity Between MAP-ELA and MAP-MA for 3rd Grade.

Predictor	Fall		Winter		Spring	
	ELA	*Math*	*ELA*	*Math*	*ELA*	*Math*
WD-WW	.33	-.01	.50*	.09	.45	.50*
WD-WSC	.43	.16	.43	.22	.44	.51*
WD-CLS	.35	.06	.44	.16	.43	.51*

PW-WW	.48*	-.03	.54*	.05	.65**	.42
PW-WSC	.50*	.03	.56*	.12	.66**	.46*
PW-CWS	.44	-.05	.46*	.05	.67**	.53*
SAEBRS-AB	.45	.39	.62**	.63**	.63**	.53*

Summary. PW had consistently higher correlations with the MAP-ELA (r = .44-.67) than WD (r = .33-.50) across time-points. Although both PW and WD metrics were had higher correlations with the MAP-ELA than the MAP-MA in fall and winter, only PW had higher correlations with the MAP-ELA than the MAP-MA (r = .65-.67 and r = .46-.53, respectively) in the spring. Therefore, PW appeared to be the most robust measure of MAP-ELA performance across time-points. SAEBRS-AB correlated as well or better than either WD or PW with the MAP-ELA across time-points. However, divergent validity indicated that both PW and WD were more discriminative than SAEBRS-AB.

Sensitivity to Growth

Sensitivity to growth is first evaluated by examining descriptive statistics (mean and SD) across time-points within and between grades for each predictor. For the descriptive analyses, only CBM-W metrics including accuracy (WSC, CLS, CWS) will be analyzed because they were consistently more reliable and valid across previous analyses within this study. Also, the best metric for each CBM-W, as determined by previous reliability and validity analyses, was examined using a two-way mixed ANOVA to determine if there were any statistically significant differences between time-points and grades.

Descriptive Statistics Across Time-Points. Means and standard deviations of WD-WSC and CLS, the most reliable and valid metrics according to previous analysis,

for each grade at each time-point are reported in Table 27. Means increased for each metric across progressive time-points (fall-winter-spring) within each grade as well as across grades for each respective time-point (e.g., 2nd graders performed higher than 1st graders in fall). Descriptive analysis indicates WD-WSC and CLS reflect growth across time and grades.

Table 27

Mean & Standard Deviation of WD-WSC & WD-CLS Across Grades/Times

	WSC			CLS		
	Fall	Winter	Spring	Fall	Winter	Spring
1st	5 (4.5)	6 (5.0)	8 (6.5)	49 (27.6)	57 (29.2)	67 (36.1)
2nd	12 (10.1)	15 (10.5)	19 (12.6)	86 (49.4)	105 (51.2)	123 (62.7)
3rd	15 (7.9)	17 (8.4)	21 (9.8)	103 (39.6)	121 (40.4)	138 (39.2)

Means and standard deviations of PW-WSC and CWS, the metrics incorporating accuracy, for each grade at each time-point are reported in Table 28. Means increased for each metric across progressive time-points (fall-winter-spring) within each grade as well as across grades for each respective time-point (e.g., 2nd graders performed higher than 1st graders in fall) except for fall between 2nd and 3rd grade for both PW-WSC and CWS, for which 2nd grade performed better than 3rd grade. Descriptive analysis indicates PW-WSC and CWS consistently reflected growth across time-points within grades but differences between 2nd and 3rd grade performance were inconsistent.

Table 28

Mean & Standard Deviation of PW-WSC & PW-CWS Across Grades/Times

	WSC			CWS		
	Fall	Winter	Spring	Fall	Winter	Spring
1st	11(5.0)	14(5.1)	18(7.3)	9(4.8)	11(5.8)	17(9.9)

2nd	27(13.8)	29(12.8)	34(14.1)	24(10.8)	27(9.9)	35(15.7)
3rd	25(12.9)	30(9.8)	35(13.3)	23(13.6)	29(11.3)	36(16.0)

For SAEBRS-AB, means and SD at successive time-points are provided in Table 13. Evidence of sensitivity to growth over time is not clear but this does not mean that SAEBRS-AB is not sensitive to change. SAEBRS does not assume that students are receiving instruction in the relevant skills and should therefore be growing or improving across time. Indeed, it may actually be the case that many students with behavior and motivation needs are receiving little to no instruction in these skills and may therefore not consistently grow across time in these skills. Descriptive statistics indicated that further analyses regarding sensitivity to growth (ANOVA) for SAEBRS-AB were not warranted.

Statistical Analyses of Sensitivity to Growth. *Word Dictation.* Statistically significant differences between grades and time-points for average WD-WSC were evaluated using a two-way mixed ANOVA with the between-subjects factor of grade (1st, 2nd, 3rd) and the within-subjects factor of season (fall, winter, spring). There were no outliers, as assessed by examination of studentized residuals (i.e., a residual measured in standard units; Field, 2015) for values greater than ±3. The variables were normally distributed, as assessed by a Normal Q-Q Plot. There was not homogeneity of variances, as assessed by Levene's test of homogeneity of variance ($p < .05$). Thus, the dependent variable (Average WSC) was transformed using Square Root. Again, there were no outliers, as assessed by examination of studentized residuals for values greater than ±3 and the variables were normally distributed, as assessed by a Normal Q-Q Plot. After transformation, there was homogeneity of variances, as assessed by Levene's test of homogeneity of variance ($p > .05$). There was homogeneity of covariance, as assessed by

Box's test of equality of covariance matrices (p = .079). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(2) = 27.471$, p = .000. Maxwell & Delaney (2004) suggest using the Greenhouse-Geisser correction, especially if estimated epsilon (ε) is less than 0.75. Greenhouse-Geisser was used for further interpretation as the Mauchly's test of sphericity indicated ε = .726.

There was no statistically significant interaction between grade and time-point on WD-WSC performance, $F(2.904, 85.677) = 1.004$, p = .393, partial η^2 = .033, ε = .726. The main effect of grade showed that there was a statistically significant difference in mean WD-WSC between grades $F(2, 59) = 48.104$, $p < .001$, partial η^2 = .280. Follow-up pairwise comparisons were run for each simple main effect with reported 95% confidence intervals and p-values Bonferroni-adjusted within each simple main effect. Pairwise comparisons indicated the marginal means for 1st grade WD-WSC were 2.285 (SE = .271) and 3.487 (SE = .258) for 2nd grade, a statistically significant mean difference of 1.202, 95% CI [-2.124, -.280], p = .006. Third grade marginal means for WD-WSC were 4.031 (SE = .252), also a statistically significant mean difference with 1st grade WD-WSC of 1.746, 95% CI [-2.659, .-834], $p < .001$. The marginal means for 2nd grade WD-WSC were 3.487 (SE = .258) and 3rd grade 4.031 (SE = .252), not a statistically significant mean difference of .545, 95% CI [-1.433, .344], p = .409.

Tests of within-subjects effects showed a statistically significant difference in mean WD-WSC between seasons $F(1.452, 85.677) = 11.029$, $p < .001$, partial η^2 = .480. Follow-up pairwise comparisons were run for each simple main effect with reported 95% confidence intervals and p-values Bonferroni-adjusted within each simple main effect. Pairwise comparisons indicated the marginal means for winter WD-WSC were 3.278 (SE

= .151) and 2.902 (*SE* = .155) for fall, a statistically significant mean difference of .376, 95% CI [.231, -.522], $p < .001$. Marginal means for spring WD-WSC were 3.622 (*SE* = .161) for a statistically significant mean difference with winter of .344, 95% CI [.206, .481], $p < .001$.

The differences between time-points on WSC for WD were also analyzed using one-way RM-ANOVA at each grade, using non-transformed data for easier interpretation. For 1st grade WD-WSC, there was one moderate outlier in winter, as determined by boxplots for values greater than 1.5 box-lengths from the edge of the box. WD-WSC was not normally distributed across time-points for 1st grade, as assessed by Shapiro-Wilk's test ($p < .05$). Although the assumption of normal distribution was violated, analysis was continued because one-way RM-ANOVA is robust to deviations from normality. WD-WSC increased from fall (*M* = 5.03, *SD* = 4.68) to winter (*M*= 6.34, *SD* = 5.07) and again to spring (*M* = 7.95*, SD* = 6.58). Mauchly's test of sphericity indicated that the assumption of sphericity had been violated $\chi^2(2) = 14.697$, $p = .001$. Epsilon (ε) was 0.633, as calculated according to Greenhouse & Geisser (1959), and was used to correct the one-way RM-ANOVA. WD-WSC was statistically significantly different at the different time points, $F(1.267, 22.803) = 14.257$, $p < .001$, partial η^2 = .442, partial ω^2 = .317. There was an increase in WD-WSC from fall (*M* = 5.03, *SD* = 4.68) to winter WSC (*M*= 6.34, *SD* = 5.07), a statistically significant mean increase of 1.316, 95% CI [0.28, 2.35], $p = .011$ and an increase in WD-WSC from winter WD-WSC to spring (*M* = 7.95*, SD* = 6.58), a statistically significant mean increase of 1.61, 95% CI [0.36, 2.85], $p = .009$, for 1st grade.

For 2nd grade WD-WSC, there were no outliers as determined by boxplots for values greater than 1.5 box-lengths from the edge of the box. WD-WSC was not normally distributed in fall or winter for 2nd grade, as assessed by Shapiro-Wilk's test ($p < .05$). Again, analysis was continued because one-way RM- ANOVA is robust to deviations from normality. WD-WSC increased from fall ($M = 11.29$, $SD = 9.45$) to winter ($M = 14.33$, $SD = 10.41$) and again to spring ($M = 17.19$, $SD = 11.65$) for 2nd grade. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated $\chi^2(2) = 10.885$, $p = .004$. Epsilon (ε) was 0.696, as calculated according to Greenhouse & Geisser (1959), and was used to correct the one-way RM-ANOVA. WD-WSC was statistically significantly different at the different time points, $F(1.393, 27.853) = 13.726$, $p < .001$, partial $\eta^2 = .407$, partial $\omega^2 = .288$. There was an increase in WD-WSC from fall ($M = 11.29$, $SD = 9.45$) to winter WSC ($M = 14.33$, $SD = 10.41$), a statistically significant mean increase of 3.05, 95% CI 0.60, 5.49], $p = .012$ and an increase in WD-WSC from winter to spring ($M = 17.19$, $SD = 11.65$), a statistically significant mean increase of 2.86, 95% CI [0.48, 5.23], $p = .015$, for 2nd grade.

For 3rd grade WD-WSC, there were no outliers. WD-WSC was normally distributed across time-points, as assessed by Shapiro-Wilk's test ($p > .05$). Mauchly's test of sphericity indicated that the assumption of sphericity had been violated $\chi^2(2) = 8.019$, $p = .018$. Epsilon (ε) was 0.752, as calculated according to Greenhouse & Geisser (1959), and was used to correct the one-way RM-ANOVA. WD-WSC was statistically significantly different at the different time points, $F(1.503, 31.571) = 23.468$, $p < .001$, partial $\eta^2 = .528$, partial $\omega^2 = .405$. There was an increase in WD-WSC from fall ($M = 14.64$, $SD = 7.68$) to winter WSC ($M = 17.00$, $SD = 8.16$), a statistically significant mean

increase of 2.36, 95% CI [0.64, 4.09], $p = .006$ and an increase in WSC from winter to spring ($M = 20.71$, $SD = 2.09$), a statistically significant mean increase of 3.71, 95% CI [1.49, 5.92], $p = .001$, for 3rd grade.

Statistically significant differences between grades and time points for average WD-CLS were also evaluated using a two-way mixed ANOVA. There were no outliers, as assessed by examination of studentized residuals for values greater than ±3. The variables were normally distributed, as assess by Normal Q-Q Plot. There was not homogeneity of variances for winter, as assessed by Levene's test of homogeneity of variance ($p < .05$) but there was homogeneity of variance for fall and spring ($p > .05$). Thus, the dependent variable (average WD-CLS) was transformed using Square Root. Again, there were no outliers and the variables were normally distributed. After transformation, there was homogeneity of variances, as assessed by Levene's test of homogeneity of variance ($p > .05$). There was homogeneity of covariance, as assessed by Box's test of equality of covariance matrices ($p = .860$). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(2) = 27.275$, $p = .000$. Greenhouse-Geisser was used for further interpretation as the Mauchly's test of sphericity indicate $\varepsilon = .745$.

There was no statistically significant interaction between grade and time on WD-CLS performance, $F(2.981, 87.929) = .490$, $p = .689$, partial $\eta^2 = .016$, $\varepsilon = .745$. The main effect of grade showed that there was a statistically significant difference in mean WD-CLS between grades $F(2, 59) = 12.876$, $p < .001$, partial $\eta^2 = .304$, partial $\omega^2 = .288$. Follow-up pairwise comparisons were run for each simple main effect with reported 95% confidence intervals and p-values Bonferroni-adjusted within each simple main effect.

Pairwise comparisons indicated the marginal means for 1st grade WD-CLS were 7.379 (*SE* = .502) and 9.648 (*SE* = .478) for 2nd grade, a statistically significant mean difference of 2.269, 95% CI [.561, 3.977], $p = .005$. Third grade marginal means for WD-CLS were 10.824 (*SE* = .467), also a statistically significant mean difference with 1st grade CLS of 3.445, 95% CI [1.755, 5.134], $p < .001$. The marginal means difference between 3rd and 2nd grade WD-CLS was not statistically significant, 1.176, 95% CI [-.470, 2.822], $p = .250$.

Tests of within-subjects effects showed a statistically significant difference in mean WD-CLS between seasons $F(1.490, 87.929) = 34.179$, $p < .001$, partial $\eta^2 = .367$, partial $\omega^2 = .288$. Follow-up pairwise comparisons were run for each simple main effect of time with reported 95% confidence intervals and *p*-values Bonferroni-adjusted within each simple main effect. Pairwise comparisons indicated the marginal means for winter WD-CLS were 9.295 (*SE* = .283) and 8.589 (*SE* = .296) for fall, a statistically significant mean difference of .707, 95% CI [.333, 1.080], $p < .001$. Marginal means for spring WD-CLS were 9.967 (*SE* = .304) for a statistically significant mean difference with winter of .672, 95% CI [.353, .990], $p < .001$.

The differences between time-points on CLS for WD were analyzed using one-way RM-ANOVA at each grade, using non-transformed data. For 1st grade WD-CLS, there were no outliers and WD-CLS was normally distributed across time-points, as assessed by Shapiro-Wilk's test ($p > .05$). WD-CLS increased from fall ($M = 48.95$, $SD = 29.54$) to winter ($M = 57.58$, $SD = 30.49$) and again to spring ($M = 68.45$, $SD = 36.63$). Mauchly's test of sphericity indicated that the assumption of sphericity was met $\chi^2(2) = 2.792$, $p = .248$. WD-CLS was statistically significantly different at the different time

points, $F(2, 36) = 17.174$, $p < .001$, partial $\eta^2 = .488$, partial $\omega^2 = .362$. There was an increase in WD-CLS from fall ($M = 48.95$, $SD = 29.54$) to winter ($M = 57.58$, $SD = 30.49$), a statistically significant mean increase of 8.632, 95% CI [1.35, 15.92], $p = .017$ and an increase in WD-CLS from winter to spring ($M = 68.45$, $SD = 36.63$), a statistically significant mean increase of 10.87, 95% CI [2.25, 19.49], $p = .011$ for 1st grade.

For 2nd grade WD-CLS, there were no outliers and WD-CLS was normally distributed in fall and winter, as assessed by Shapiro-Wilk's test ($p > .05$), but not spring ($p < .05$). Although the assumption of normal distribution was violated in spring, analysis was continued because one-way RM-ANOVA is robust to deviations from normality. WD-CLS increased from fall ($M = 82.05$, $SD = 47.63$) to winter ($M = 102.10$, $SD = 53.15$) and again to spring ($M = 114.60$, $SD = 59.81$) for 2nd grade. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated $\chi^2(2) = 7.988$, $p = .018$. Epsilon (ε) was 0.744, as calculated according to Greenhouse & Geisser (1959), and was used to correct the one-way RM-ANOVA. WD-CLS was statistically significantly different at the different time points, $F(1.489, 29.779) = 13.137$, $p < .001$, partial $\eta^2 = .396$, partial $\omega^2 = .278$. There was an increase in WD-CLS from fall ($M = 82.05$, $SD = 47.63$) to winter ($M = 102.10$, $SD = 53.15$), a statistically significant mean increase of 20.05, 95% CI [4.83, 35.27], $p = .008$ and an increase in WD-CLS from winter to spring ($M = 114.60$, $SD = 59.81$), which was not a statistically significant mean increase of 4.97, 95% CI [-0.49, 25.49], $p = .062$, for 2nd grade.

For 3rd grade WD-CLS, there was one moderate outlier in fall and WD-CLS was normally distributed across time-points, as assessed by Shapiro-Wilk's test ($p > .05$).

Mauchly's test of sphericity indicated that the assumption of sphericity was met $\chi^2(2) = 3.424$, $p = .181$. WD-CLS was statistically significantly different at the different time points, $F(2, 42) = 17.205$, $p < .001$, partial $\eta^2 = .450$, partial $\omega^2 = .329$. There was an increase in WD-CLS from fall ($M = 103.52$, $SD = 8.20$) to winter ($M = 120.00$, $SD = 8.60$), a statistically significant mean increase of 16.48, 95% CI [2.38, 30.57], $p = .019$ and an increase in WD-CLS from winter to spring ($M = 138.11$, $SD = 10.50$), a statistically significant mean increase of 18.11, 95% CI [4.72, 31.51], $p = .006$, for 3rd grade.

In summary, descriptive statistics indicated that both WD-WSC and CLS were sensitive to growth within and across grades. Statistical analysis using ANOVA indicated significant growth across time-points within each grade for both WD-WSC and CLS, with the exception of winter to spring average WD-CLS for 2nd grade. ANOVA and Post Hoc tests indicated that both WD-WSC and CLS significantly discriminated between 1st and 2nd and 1st and 3rd grade participants but not between 2nd and 3rd grade participants.

Picture Word. PW-WSC was also evaluated using two-way mixed model ANOVAs. There was not homogeneity of variances, as assessed by Levene's test of homogeneity of variance ($p < .05$). Thus, the dependent variable (Average PW-WSC) was transformed using Square Root. There were no outliers and the variables were normally distributed. After transformation, there was homogeneity of variances, as assessed by Levene's test of homogeneity of variance ($p > .05$). There was no homogeneity of covariance, as assessed by Box's test of equality of covariance matrices ($p = .039$), but analyses continued despite this violation. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(2)$

= 39.839, $p < .001$. Greenhouse-Geisser was used for further interpretation as the Mauchly's test of sphericity indicate $\varepsilon = .668$.

There was no statistically significant interaction between grade and time-point on PW-WSC performance, $F(2.672, 78.832) = 1.156$, $p = .329$, partial $\eta^2 = .038$, $\varepsilon = .668$. The main effect of grade showed that there was a statistically significant difference in mean PW-WSC between grades $F(2, 59) = 52.438$, $p < .001$, partial $\eta^2 = .381$. Follow-up pairwise comparisons were run for each simple main effect of grade with reported 95% confidence intervals and *p*-values Bonferroni-adjusted within each simple main effect. Pairwise comparisons indicated the marginal means for 1st grade PW-WSC were 3.687 ($SE = .231$) and 5.358 ($SE = .205$) for 2nd grade, a statistically significant mean difference of 1.671, 95% CI [-2.431, -.910], $p < .001$. Third grade marginal means for PW-WSC were 5.323 ($SE = .214$), also a statistically significant mean difference with 1st grade WSC of 1.636, 95% CI [-2.412, .-859], $p < .001$. The marginal means for 2nd grade PW-WSC were 5.358 ($SE = .205$) and 3rd grade 5.323 ($SE = .214$), not a statistically significant mean difference of .035, 95% CI [-.695, .765], $p = 1.000$.

Because of violations of sphericity and lack of homogeneity of covariance, tests of within-subjects effects for time-point were not evaluated using the two-way mixed model ANOVA, but three separate one-way RM-ANOVA were used, one for each grade with the within-subjects factor set as fall, winter, and spring. Also, non-transformed data was used for ease of interpretation. For 1st grade PW-WSC, there were no outliers as determined by boxplots for values greater than 1.5 box-lengths from the edge of the box. PW-WSC was normally distributed across time-points for 1st grade, as assessed by Shapiro-Wilk's test ($p > .05$). PW-WSC increased from fall ($M = 10.75$, $SD = 5.04$) to

winter (*M*= 13.61, *SD* = 4.77) and again to spring (*M = 18.78, SD* = 7.39). Mauchly's test of sphericity indicated that the assumption of sphericity had been violated $\chi^2(2) = 9.127$, $p = .010$. Epsilon (ε) was 0.697, as calculated according to Greenhouse & Geisser (1959), and was used to correct the one-way RM-ANOVA. PW-WSC was statistically significantly different at the different time points, $F(1.394, 23.698) = 35.878$, $p < .001$, partial $\eta^2 = .679$, partial $\omega^2 = .564$. There was an increase in PW-WSC from fall (*M* = 10.75, *SD* = 1.19) to winter WSC (*M* = 13.61, *SD* = 1.12), a statistically significant mean increase of 2.86, 95% CI [1.14, 4.58], $p < .001$ and an increase in PW-WSC from winter to spring (*M* = 18.78, *SD* = 1.74), a statistically significant mean increase of 5.17, 95% CI [2.04, 8.29], $p < .001$, for 1st grade.

For 2nd grade PW-WSC, there were was one moderate and one extreme outlier and PW-WSC was not normally distributed across time-points, as assessed by Shapiro-Wilk's test ($p < .05$). Although the assumption of normal distribution was violated, analysis was continued because one-way RM-ANOVA is robust to deviations from normality and could be run to maintain consistency of procedures across analyses. PW-WSC increased from fall (*M* = 27.44, *SD* = 13.77) to winter (*M*= 29.26, *SD* = 13.17) and again to spring (*M = 33.85, SD* = 14.20) for 2nd grade. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated $\chi^2(2) = 13.237$, $p = .001$. Epsilon (ε) was 0.681, as calculated according to Greenhouse & Geisser (1959), and was used to correct the one-way RM-ANOVA. PW-WSC was statistically significantly different at the different time points, $F(1.363, 29.981) = 9.643$, $p = .002$, partial $\eta^2 = .305$, partial $\omega^2 = .200$. There was an increase in PW-WSC from fall (*M* = 27.44, *SD* = 13.77) to winter (*M*= 29.26, *SD* = 13.17), not a statistically significant mean increase of

1.83, 95% CI [-0.91, 4.56], p = .198 and an increase in PW-WSC from winter to spring (*M = 33.85, SD* = 14.20), a statistically significant mean increase of 4.59, 95% CI [.16, 9.02], p =.02, for 2nd grade.

For 3rd grade PW-WSC, there were was one moderate outlier in the fall and PW-WSC was normally distributed for fall and winter, as assessed by Shapiro-Wilk's test (p > .05). However, WSC was not normally distributed for spring in 3rd grade, as assessed by Shapiro-Wilk's test (p < .05). Although the assumption of normality was violated, analysis was continued. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated $\chi^2(2) = 21.879$, $p < .001$. Epsilon (ε) was 0.594, as calculated according to Greenhouse & Geisser (1959), and was used to correct the one-way RM-ANOVA. PW-WSC was statistically significantly different at the different time points, $F(1.188, 23.755) = 9.285$, $p < .004$, partial $\eta^2 = .317$, partial $\omega^2 = .208$. There was an increase in PW-WSC from fall (*M* = 24.93, *SD* = 12.98) to winter (*M*= 29.21, *SD* = 10.14), a statistically significant mean increase of 4.29, 95% CI [-0.16, 8.74], p = .032 and an increase in PW-WSC from winter to spring (*M = 35.17, SD* = 13.65), a statistically significant mean increase of 5.95, 95% CI [-0.29, 12.20], p = .034, for 3rd grade.

Statistically significant differences between grades and time points for PW-CWS were also evaluated using a two-way mixed ANOVA. There was not homogeneity of variances for fall or spring, as assessed by Levene's test of homogeneity of variance (p < .05) but there was homogeneity of variance for winter (p > .05). Thus, the dependent variable (PW-CWS) was transformed using Square Root. PW-CWS was normally distributed as assessed by Shapiro-Wilk's test (p > .05). There were no outliers and the

variables were normally distributed. After transformation, there was homogeneity of variances, as assessed by Levene's test of homogeneity of variance ($p > .05$). There was homogeneity of covariance, as assessed by Box's test of equality of covariance matrices ($p = .484$). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi^2(2) = 43.850$, $p < .001$. Greenhouse-Geisser was used for further interpretation as the Mauchly's test of sphericity indicate $\varepsilon = .653$.

There was no statistically significant interaction between grade and time on PW-CWS performance, $F(2.614, 77.100) = .632$, $p = .575$, partial $\eta^2 = .021$, $\varepsilon = .653$. The main effect of grade showed that there was a statistically significant difference in mean PW-CWS between grades $F(2, 59) = 20.956$, $p < .001$, partial $\eta^2 = .415$. Follow-up pairwise comparisons were run for each simple main effect with reported 95% confidence intervals and *p*-values Bonferroni-adjusted within each simple main effect. Pairwise comparisons indicated the marginal means for 1st grade PW-CWS were 3.255 ($SE = .249$) and 5.151 ($SE = .221$) for 2nd grade, a statistically significant mean difference of 1.896, 95% CI 1.075, 2.717], $p < .001$. Third grade marginal means for PW-CWS were 5.194 ($SE = .231$), also a statistically significant mean difference with 1st grade CWS of 1.939, 95% CI [1.101, 2.776], $p < .001$. The marginal means difference between 3rd and 2nd grade PW-CWS was not statistically significant, .043, 95% CI -0.744, 0.830], $p = 1.00$.

Tests of within-subjects effects showed a statistically significant difference in mean PW-CWS between seasons $F(1.307, 77.100) = 40.056$, $p < .001$, partial $\eta^2 = .404$. Follow-up analyses were conducted using RM-ANOVA for each grade using non-transformed average PW-CWS. For 1st grade PW-CWS, there were no outliers and PW-CWS was normally distributed across time-points, as assessed by Shapiro-Wilk's test (*p*

> .05). Mauchly's test of sphericity indicated that the assumption of sphericity had been violated $\chi^2(2) = 11.717$, $p = .003$. Epsilon (ε) was 0.658, as calculated according to Greenhouse & Geisser (1959), and was used to correct the one-way RM-ANOVA. PW-CWS was statistically significantly different at the different time points, $F(1.316, 22.380) = 22.823$, $p < .001$, partial $\eta^2 = .573$, partial $\omega^2 = .447$. There was an increase in PW-CWS from fall ($M = 7.944$, $SD = 4.76$) to winter ($M = 10.833$, $SD = 5.91$), a statistically significant mean increase of 2.89, 95% CI [.69, 5.09], $p = .009$ and an increase in PW-CWS from winter to spring ($M = 17.306$, $SD = 10.22$), a statistically significant mean increase of 6.47, 95% CI [2.56, 10.38], $p = .001$, for 1st grade.

For 2nd grade PW-CWS, there were three moderate outliers for fall and one for spring. PW-CWS was not normally distributed in fall for 2nd grade, as assessed by Shapiro-Wilk's test ($p < .05$), but was for winter and spring. Although the assumption of normal distribution was violated, analysis was continued. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated $\chi^2(2) = 9.284$, $p = .010$. Epsilon (ε) was 0.737, as calculated according to Greenhouse & Geisser (1959), and was used to correct the one-way RM-ANOVA. PW-CWS was statistically significantly different at the different time points, $F(1.474, 32.417) = 17.226$, $p < .001$, partial $\eta^2 = .439$, partial $\omega^2 = .320$. There was an increase in PW-CWS from fall ($M = 23.522$, $SD = 10.76$) to winter ($M = 26.326$, $SD = 10.21$), not a statistically significant mean increase of 2.80, 95% CI [-0.50, 6.12], $p = .115$ and an increase in PW-CWS from winter to spring ($M = 34.17$, $SD = 15.53$), a statistically significant mean increase of 7.85, 95% CI [2.85, 12.85], $p = .002$, for 2nd grade.

For 3rd grade PW-CWS, there were no outliers and PW-CWS was normally distributed across time-points, as assessed by Shapiro-Wilk's test ($p > .05$). Mauchly's test of sphericity indicated that the assumption of sphericity had been violated $\chi^2(2) = 27.182$, $p < .001$. Epsilon (ε) was 0.568, as calculated according to Greenhouse & Geisser (1959), and was used to correct the one-way RM-ANOVA. PW-CWS was significantly different at the different time points, $F(1.136, 22.716) = 12.769$, $p = .001$, partial $\eta^2 = .390$, partial $\omega^2 = .272$. There was an increase in PW-CWS from fall ($M = 22.36$, $SD = 13.56$) to winter ($M = 28.55$, $SD = 11.45$), a statistically significant mean increase of 6.19, 95% [CI 2.58, 9.80], $p = .001$ and an increase in PW-CWS from winter to spring ($M = 36.02$, $SD = 16.38$), a statistically significant mean increase of 7.48, 95% CI [.511, 14.44], $p = .033$, for 3rd grade.

In summary, descriptive analysis indicates PW-WSC and CWS reflect growth across time-points within grades but differences between 2nd and 3rd grade performance are inconsistent, especially in fall. This is relevant to growth in that 3rd graders should perform better than 2nd graders. Statistical analysis using ANOVA indicated significant growth across all time-points for PW-WSC and CWS in 1st grade and 3rd grade but only between winter and spring for 2nd grade. ANOVA and Post Hoc tests indicated that both PW-WSC and CWS significantly discriminated between 1st and 2nd and 1st and 3rd grade participants but not between 2nd and 3rd grade participants.

EL Performance Compared to the General Population

Word Dictation (WD). In order to examine the difference in performance on WD between ELs in this study and the general population, descriptive statistics for WD-WW, WSC, and CLS from this study were compared to those from a previous benchmarking

study drawn from the general population using the same CBM-W forms (McMaster, Brandes, Herriges, & Jung, 2014). For the benchmarking study drawn from the general population, first through third graders ($N = 274$) were selected from two elementary schools in a large urban city school district in the Midwest. The school district served 34,400 K-12 students in 71 schools. Student demographics were 32.8% White, 36.2% African American, 18.8% Hispanic, 7.6% Asian, and 4.6% American Indian. Nineteen percent of the students were receiving special education services, 21% were receiving English language services, and 65.6% were eligible for free and reduced lunch service. Further demographic information is available in via McMaster and colleagues (2014) technical report.

Scores reported for both the technical report from the general population and this study are average mean performance and standard deviations for each metric within each respective grade and for each time-point. Results are available in Table 29. Generally, ELs had lower scores across metrics, grades, and time-points except for WW in fall and spring for 2nd grade and spring for 3rd grade. The sizes of the performance differences were generally larger for WD-WSC and WD-CLS than WD-WW. EL performance fell within the respective SD of mean performance of the general population across metrics and times with the exception of WD-WSC in fall, winter, and spring as well as WD-CLS in fall and winter for 3rd grade.

Table 29

Comparison of Mean EL and General Population on WD

	Fall		Winter		Spring	
	Gen Pop	EL	Gen Pop	EL	Gen Pop	EL
1ST Grade	*N = 82*	*N = 22*	*N = 89*	*N = 21*	*N = 84*	*N = 20*
WW	19.6 (5.5)	16.7 (6.5)	21.5 (6.6)	19.8 (6.7)	24.6 (7.3)	22.6 (8.1)
WSC	9.7 (6.0)	4.9 (4.5)	12.5 (7.7)	5.9 (4.9)	16.3 (9.1)	7.7 (6.5)

CLS	68.9 (31.9)	48.6 (27.6)	83.6 (37.9)	57.1 (29.2)	101.1 (43.4)	67.2 (36.1)
2nd Grade	*N = 96*	*N = 22*	*N = 99*	*N = 24*	*N = 97*	*N = 25*
WW	25.3 (6.1)	23.3 (9.1)	23.1 (9.3)	27.4 (9.3)	27.4 (10.9)	30.7 (11.3)
WSC	17.6 (7.8)	12.2 (10.1)	23.1 (9.3)	15.2 (10.5)	27.4 (10.9)	19.1 (12.6)
CLS	108.1 (36.6)	85.6 (49.4)	134.7 (42.7)	104.8 (51.8)	155.3 (49.5)	122.5 (62.7)
3rd Grade	*N = 74*	*N = 24*	*N = 76*	*N = 24*	*N = 77*	*N= 22*
WW	33.0 (11.9)	26.3 (7.6)	30.8 (13.7)	30.7 (7.4)	31.8 (12.6)	32.8 (9.2)
WSC	28.9 (14.1)	14.6 (7.9)	30.8 (13.7)	17.0 (8.4)	31.8 (12.6)	20.7 (9.8)
CLS	158.7 (67.6)	102.9 (39.6)	170.5 (64.1)	120.6 (40.4)	175.2 (58.2)	138.1 (49.2)

Picture Word (PW). In order to examine the difference in performance on PW between ELs in this study and the general population, descriptive statistics for PW-WW, WSC, and CWS from this study were compared to those from a previous benchmarking study drawn from the general population (Allen et al., n.d.). The study by Allen and colleagues (n.d.) included 612 students in grades 1-3 for PW from two districts in two Midwestern states collected across the 2013-2014 and 2014-2015 academic years. Additional demographic data is available in Allen et al. (2014). Scores reported for both the study from the general population and this study are average mean performance and standard deviations for each metric within each respective grade and for each time-point. Results are available in Table 30. As with WD, ELs often scored lower than the general population across metrics and times with the exception of 2nd grade ELs outperforming the general population across metrics in the fall. Although 2nd grade ELs grew across time-points, the general population grew faster than ELs and outperformed them in winter and spring. EL performance was within their respective SD of mean performance of the general population with the exceptions of PW-WSC and CWS in winter for 1st grade.

Table 30

Comparison of Mean EL and General Population on PW

	Fall	Winter	Spring

	Gen Pop	EL	Gen Pop	EL	Gen Pop	EL
1ST Grade	*N = 80*	*N = 21*	*N = 87*	*N = 23*	*N = 73*	*N = 20*
WW	17.5 (8.5)	14.6 (6.3)	22.8 (105)	17.2 (6.6)	26.7 (10.9)	22.0 (7.7)
WSC	14.4 (8.0)	11.4 (5.0)	20.2 (10.4)	13.9 (5.1)	23.9 (10.5)	18.2 (7.3)
CWS	12.9 (8.4)	8.5 (4.8)	18.5 (12.0)	11.1 (5.8)	23.4 (12.4)	16.7 (9.9)
2nd Grade	*N = 93*	*N = 23*	*N = 94*	*N = 25*	*N = 90*	*N = 25*
WW	23.1 (8.1)	29.7 (14.6)	32.3 (11.4)	31.9 (14.3)	39.3 (11.4)	36.5 (15.1)
WSC	20.9 (9.8)	27.4 (13.8)	30.2 (11.4)	29.3 (12.8)	37.2 (11.6)	34.3 (14.1)
CWS	20.7 (9.8)	23.5 (10.8)	31.5 (13.6)	26.7 (9.9)	39.2 (14.4)	35.2 (15.7)
3rd Grade	*N = 70*	*N = 23*	*N = 78*	*N = 23*	*N = 72*	*N= 22*
WW	28.0 (10.6)	27.7 (13.6)	38.9 (12.6)	32.0 (10.2)	39.0 (14.1)	37.9 (13.2)
WSC	26.5 (10.9)	24.9 (12.9)	37.3 (12.7)	29.6 (9.8)	37.3 (13.9)	35.3 (13.3)
CWS	27.7 (12.9)	22.6 (13.6)	39.4 (14.7)	29.2 (11.3)	38.9 (15.8)	35.8 (16.0)

Summary. ELs generally scored slightly below the general population but mean performances were within a SD of each other, suggesting there were no true differences in performance between the populations. Possible exceptions include WD-WSC and WD-CLS in 3rd grade because EL performance approached a SD below that of the general population.

Integration of CBM-W and SAEBRS-AB to Predict Academic Performance

The utility of integrating CBM-W and SAEBRS-AB were explored using two different approaches across both criterion measures. First, hierarchical regression was used to determine how much additional variance in performance on the respective criterion measure was explained over and beyond that accounted for by WD and PW. Next, logistic regression and ROC curve analyses were used to determine to what extent combining SAEBRS-AB with each CBM-W improved diagnostic accuracy with each criterion measure.

Hierarchical Regression: ACCESS-W. *Word Dictation.* First, hierarchical regression was used to determine how much variance was explained in ACCESS-W performance with the addition of SAEBRS-AB in winter. A 4-step model was used for winter predictors with the first step being control demographics (i.e., grade), the second

step included the ACCESS-ORC to control for English oral language proficiency, the third step added in average WD-CLS in winter, and the final step added the winter SAEBRS-AB score. The addition of ACCESS-ORC as a second step was intended to address the question of whether or not English oral proficiency explained a significant amount of variance in ACCESS-W performance. A lack of initial English oral proficiency data from the previous school year did not allow for previous analyses or predictive analyses to include English oral proficiency in interpreting results, but these analyses will be used to determine how critical this missing information was. For the third step, WD-CLS was chosen as the best metric because it performed equitably with WD-WSC across analyses and WD-CLS was less prone to skew and floor effects than WD-WSC.

A linear relationship existed between the ACCESS-W and the independent variables collectively, as evidenced by a scatterplot of studentized residuals and unstandardized predicted values. Linear relationships were also evidenced in partial regression plots of each independent variable with the dependent variable. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. For multicolinearity, no independent variable was correlated at .70 or above with each other and VIF statistics were all less than 10. There were no studentized deleted residuals greater than ±3 standard deviations, no leverage values greater than 0.2, or values for Cook's distance above 1. The assumption of normality was met, as assessed by a Q-Q Plot.

The addition of ACCESS-ORC in winter to the prediction of ACCESS-W performance (Model 2), which was used to explore the importance of including oral

English proficiency when examining predictive validity, led to a statistically significant increase in R^2 of .183, $F(1, 64) = 23.088$, $p < .001$. The addition of WD-CLS to the prediction of ACCESS-W performance (Model 3) led to a statistically significant increase in R^2 of .204, $F(1, 63) = 42.283$, $p < .001$. The addition of SAEBRS-AB in winter to the prediction of ACCESS-W performance (Model 4) led to an increase in R^2 of .013, $F(1, 62) = 2.682$, not statistically significant $p = .107$. The full model of grade, ACCESS-ORC, WD-CLS in winter, and winter SAEBRS-AB to predict ACCESS-W performance (Model 4) was statistically significant, $R^2 = .690$, $F(4, 62) = 37.767$, $p < .001$, adjusted $R^2 = .690$. Summary results for each model are available in Table 31.

Table 31

Hierarchical Regression Predicting ACCESS-W from Winter WD-CLS

	Writing ACCESS							
	Model 1		Model 2		Model 3		Model 4	
Variable	B	β	B	β	B	β	B	β
Constant	243.3	-	162.0		183.3		169.3	
Grade	22.3***	.56	23.5***	.59	11.3**	.28	13.3***	.33
ORC	-	-	.26***	.43	.2**	.26	.1**	.25
Win-CLS	-	-	-	-	.4***	.6	.3***	.52
Win-AB	-	-	-	-	-	-	1.1	.12
R^2	.31		.49		.70		.71	
F	29.16***		31.08***		48.18***		37.77***	
ΔR^2	-		.183		.204		.01	
ΔF	-		23.09***		42.28***		2.68	

*Note: * = p < .05; ** = p < .01; *** = p < .001; ORC = oral language composite of ACCESS*

Hierarchical regression was also used to examine variance explained in ACCESS-W from the fall predictors using a 3-step model. For model 1, grade was the independent variable, model 2 added average fall WD-CLS score, and model 3 added the fall

SAEBRS-AB score. English oral proficiency could not be added to this model because nearly 33% of the sample was missing this data for fall. A linear relationship existed between the writing subtest and the independent variables collectively, as evidenced by a scatterplot of studentized residuals and unstandardized predicted values. Linear relationships were also evidenced in partial regression plots of each independent variable with the dependent variable. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. For multicolinearity, no independent variable was correlated at .70 or above with each other and VIF statistics were all less than 10. There were no studentized deleted residuals greater than ±3 standard deviations, no leverage values greater than 0.2, or values for Cook's distance above 1. The assumption of normality was met, as assessed by a Q-Q Plot.

The addition of WD-CLS in fall to the prediction of ACCESS-W performance (Model 2) led to a statistically significant increase in R^2 of .305, F (1, 63) = 52.232, $p <$ .001. The addition of SAEBRS-AB to the prediction of ACCESS-W performance (Model 3) led to an increase in R^2 of .012, F (1, 62) = 2.146, p = .148 that was not a statistically significant increase. The full model of grade, WD-CLS in fall, and fall SAEBRS-AB to predict ACCESS-W performance (Model 3) was statistically significant, R^2 = .644, F(3, 62) = 37.364, $p <$ .001, adjusted R^2 = .627. Summary results for each model are available in Table 32.

Table 32

Hierarchical Regression Predicting ACCESS-W from Fall WD-CLS

	Writing ACCESS		
	Model 1	Model 2	Model 3

Variable	B	β	B	β	B	β
Constant	241.5***		231.3***		216.4***	
Grade	22.9***	.57	10.0**	.3	11.7**	.3
Fall-CLS	-	-	.5***	.6	.4***	.6
Fall-AB	-	-	-	-	1.2	.1
R^2	.326		.632		.644	
F	30.96***		53.99***		37.36***	
ΔR^2	-		.305		.012	
ΔF	-		52.23***		2.15	

*Note: * = $p < .05$; ** = $p < .01$; *** = $p < .001$; AB = academic behavior subscale of SAEBRS*

Picture Word. First, hierarchical regression was used to determine how much variance was explained in ACCESS-W performance with the addition of SAEBRS-AB in winter. A 4-step model was used for winter predictors as conducted with WD-CLS. For the third step, PW-WSC was selected over PW-CWS because of unacceptable alternate form reliability in fall for PW-CWS in 1st grade ($r < .70$) and relatively equitable correlations between the metrics across grades.

All assumptions were tested as with the WD-CLS and were met. The addition of ACCESS-ORC in winter to the prediction of ACCESS-W performance (Model 2) led to a statistically significant increase in R^2 of .191, $F(2, 65) = 24.585$, $p < .001$. The addition of PW-WSC to the prediction of ACCESS-W performance (Model 3) led to a statistically significant increase in R^2 of .136, $F(3, 64) = 23.597$, $p < .001$. The addition of SAEBRS-AB in winter to the prediction of ACCESS-W performance (Model 4) led to a statistically significant increase in R^2 of .028 $F(4, 63) = 5.232$, $p = .026$. The full model of grade, ACCESS-ORC, PW-WSC in winter, and winter SAEBRS-AB to predict ACCESS-W performance (Model 4) was statistically significant, $R^2 = .660$, $F(4, 63) = 30.575$, $p < .001$, adjusted $R^2 = .638$. Summary results for each model are available in Table 33.

Table 33

Hierarchical Regression Predicting ACCESS-W From Winter PW-WSC

	Writing ACCESS							
	Model 1		Model 2		Model 3		Model 4	
Variable	B	β	B	β	B	β	B	β
Constant	243.8		161.1		156.6		139.0	
Grade	22.3***	.55	23.5***	.58	14.7***	.37	17.1***	.42
ORC	-	-	.26***	.44	.25***	.41	.22***	.37
Win-WSC	-	-	-	-	1.1***	.43	1.0***	.39
Win-AB	-	-	-	-	-	-	1.6*	.18
R^2	.31		.50		.63		.66	
F	29.02***		31.99***		36.61***		30.58***	
ΔR^2	-		.191		.136		.028	
ΔF	-		24.59***		23.60***		5.23*	

*Note: * = p < .05; ** = p < .01; *** = p < .001*; ORC = oral language composite of ACCESS; AB = academic behavior subscale of SAEBRS

Hierarchical regression was also used to examine the variance in ACCESS-W performance explained by fall predictors using a 3-step model as done with WD-CLS. All assumptions were tested and met as with WD-CLS except that one leverage value was greater than 0.2 (.207), but no values for Cook's distance above 1, so analyses were continued. The addition of PW-WSC in fall to the prediction of ACCESS-W performance (Model 2) led to a statistically significant increase in R^2 of .200, $F(1, 63) = 27.397$, $p <$ .001. The addition of SAEBRS-AB in fall to the prediction of ACCESS-W performance (Model 3) led to a statistically significant increase in R^2 of .038, $F(1, 62) = 5.6\text{-}3$, $p =$ .021. The full model of grade, PW-WSC in fall, and fall SAEBRS-AB to predict ACCESS-W performance (Model 3) was statistically significant, $R^2 = .577$, $F(3, 62) =$ 28.211, $p < .001$, adjusted $R^2 = .557$. Summary results for each model are available in Table 34.

Table 34

Hierarchical Regression Predicting ACCESS-W from Fall PW-WSC

	Writing ACCESS					
	Model 1		Model 2		Model 3	
Variable	B	β	B	β	B	β
Constant	239.6***		229.9***		205.5***	
Grade	23.7***	.58	15.6***	.4	17.3**	.4
WSC	-	-	1.2***	.5	1.1***	.4
AB	-	-	-	-	1.9*	.2
R^2	.338		.539		.577	
F	32.75***		36.83**		28.21***	
ΔR^2	-		.200		.038	
ΔF	-		27.40***		5.60*	

*Note: * = p < .05; ** = p < .01; *** = p < .001; AB = academic behavior subscale of SAEBRS*

Summary. In summary, oral English proficiency explained a significant amount of variance in ACCESS-W concurrent performance when controlling for grade. Both WD-CLS and PW-WSC explained a significant amount of variance in ACCESS-W. SAEBRS-AB added significantly to the models for PW in both winter and fall but not for WD.

Hierarchical Regression: MAP-ELA. Hierarchical regressions were conducted from winter predictors to MAP-ELA and fall predictors to MAP-ELA in a parallel fashion to those run for ACCESS-W with the exception that grade was not entered as a control variable in model 1 because only 3rd grade participants completed the MAP-ELA. Regressions were run in winter to examine how much variance the addition of oral English proficiency added to the overall model (ACCESS-ORC).

Word Dictation. A linear relationship existed between the MAP-ELA and the independent variables collectively, as evidenced by a scatterplot of studentized residuals and unstandardized predicted values. Linear relationships were also evidenced in partial regression plots of each independent variable with the dependent variable. There was

homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. For multicolinearity, no independent variable was correlated at .70 or above with each other and VIF statistics were all less than 10. There were no studentized deleted residuals greater than ±3 standard deviations, five subjects had a leverage value greater than 0.2, but no values for Cook's distance above 1. All participants were retained for further analysis. The assumption of normality was met, as assessed by a Q-Q Plot.

The addition of ACCESS-ORC in winter to the prediction of MAP-ELA performance (Model 1) was not statistically significant, R^2 of .071, $F(1, 17) = 1.290$, $p = .272$. The addition of WD-CLS in winter to the prediction of MAP-ELA performance (Model 2) led to an increase in R^2 of .165, $F(2, 16) = 3.462$, $p = .081$, not statistically significant. The addition of SAEBRS-AB in winter to the prediction of MAP-ELA performance (Model 3) led to an increase in R^2 of .163 $F(3, 15) = 4.051$, $p = .062$, not statistically significant. The full model ACCESS-ORC, WD-CLS in winter, and winter SAEBRS-AB to predict MAP-ELA performance (Model 3) was statistically significant, $R^2 = .398$, $F(3, 15) = 3.311$, $p = .049$, adjusted $R^2 = .278$. Summary results for each model are available in Table 35.

Table 35

Hierarchical Regression Predicting MAP-ELA from Winter WD-CLS

	Model 1		Model 2		Model 3	
Variable	B	β	B	β	B	β
Constant	314.9		277.1		292.6	
ORC	.3	.27	.24	.21	.13	.11
Win-CLS	-	-	.5	.41	.1	.1
Win-AB	-	-	-	-	5.5	.53

R^2	.07	.24	.40
F	1.29	2.47	3.31*
ΔR^2	-	.165	.163
ΔF	-	3.462	4.051

*Note: * = $p < .05$; ** = $p < .01$; *** = $p < .001$*; ORC = oral language composite of ACCES

Hierarchical regression was also used to examine the variance in MAP-ELA performance explained by fall predictors using a 2-step model. The models were the same as those performed with the winter predictors except that the 1st step (ACCESS-ORC) was removed. All assumptions were tested and met except that two leverage values were greater than 0.2, but no values for Cook's distance above 1, so all participants were retained. The assumption of normality was met, as assessed by a Q-Q Plot.

The addition of WD-CLS in fall to the prediction of MAP-ELA performance (Model 1) was not statistically significant, $R^2 = .120$, F (1, 17) = 2.317, $p = .146$. The addition of SAEBRS-AB in fall to the prediction of MAP-ELA performance (Model 2) led to an increase in R^2 of .117, F(2, 16) = 2.446, $p = .137$, not statistically significant. The full model of WD-CLS in fall and fall SAEBRS-AB to predict MAP-ELA performance (Model 2) was not statistically significant, $R^2 = .237$, F (2, 16) = 2.480, $p = .115$, adjusted $R^2 = .141$. Summary results for each model are available in Table 36.

Table 36

Hierarchical Regression Predicting MAP-ELA from Fall WD-CLS

	Model 1		Model 2	
Variable	B	β	B	β
Constant	367.3		338.3	
Fall-CLS	.42	.35	.24	.20
Fall-AB	-	-	4.1	.37
R^2	.12		.24	
F	2.32		2.48	

ΔR^2	-	.12
ΔF	-	2.45

** = p < .05; ** = p < .01; *** = p < .001*

Picture Word. A 3-step model was conducted for average winter PW-WSC as was done with winter WD-CL. All assumptions were tested and met except that one subject had a leverage value greater than 0.2 (.26), but no values for Cook's distance above 1, so all participants were retained. The addition of ACCESS-ORC in winter to the prediction of MAP-ELA performance (Model 1) was not statistically significant, R^2 of .071, F (1, 17) = 1.290, p = .272. The addition of PW-WSC in winter to the prediction of MAP-ELA performance (Model 2) led to a statistically significant increase in R^2 of .253, F (2, 16) = 5.979, p = .026. The addition of SAEBRS-AB in winter to the prediction of ACCESS-W performance (Model 3) led to a statistically significant increase in R^2 of .175 F (3, 15) = 5.223, p = .037. The full model of ACCESS-ORC, PW-WSC in winter, and winter SAEBRS-AB to predict MAP-ELA performance (Model 3) was statistically significant, R^2 = .498, F(3, 15) = 4.964, p = .014, adjusted R^2 = .398. Summary results for each model are available in Table 37.

Table 37

Hierarchical Regression Predicting MAP-ELA from Winter PW-WSC

	Model 1		Model 2		Model 3	
Variable	B	β	B	β	B	β
Constant	314.9		306.9		297.6	
ORC	.3	.27	.14	.12	.05	.04
Win-WSC	-	-	2.1*	.52	1.4	.36
Win-AB	-	-	-	-	4.8*	.46
R^2	.07		.32		.50	
F	1.29		3.82*		4.96*	
ΔR^2	-		.253		.175	
ΔF	-		5.979*		5.223*	

*Note: * = p < .05; ** = p < .01; *** = p < .001; ORC = oral language composite of ACCESS*

Hierarchical regression was also used to examine the variance in MAP-ELA performance explained by fall predictors (PW-WSC) using a 2-step model. All assumptions were tested and met except that four leverage values were greater than 0.2, but no values for Cook's distance above 1, so all participants were maintained. The addition of PW-WSC in fall to the prediction of MAP-ELA performance (Model 1) was statistically significant, $R^2 = .25$, $F\ (1, 17) = 5.655$, $p = .029$. The addition of SAEBRS-AB in fall to the prediction of MAP-ELA performance (Model 2) led to an increase in R^2 of .111, $F(2, 16) = 2.78$, $p = .115$, not a statistically significant increase. The full model of PW-WSC in fall and fall SAEBRS-AB to predict MAP-ELA performance (Model 2) was statistically significant, $R^2 = .361$, $F(2, 16) = 4.514$, $p = .028$, adjusted $R^2 = .281$. Summary results for each model are available in Table 38.

Table 38

Hierarchical Regression Predicting ACCESS-W from Fall PW-WSC

	Model 1		Model 2	
Variable	B	β	B	β
Constant	370.6	-	332.4	-
Fall-WSC	1.6*	.5	1.4	.41
AB	-	-	3.8	.35
R^2	.25		.361	
F	5.66*		4.51*	
ΔR^2	-		.111	
ΔF	-		2.78	

** = p < .05; ** = p < .01; *** = p < .001*

Summary. In summary, oral English proficiency in winter did not explain a significant amount of variance in MAP-ELA performance for 3rd grade. WD-CLS did not

explain a significant amount of variance in MAP-ELA performance in fall or winter while PW-WSC did explain a significant amount of variance in both fall and winter. SAEBRS-AB added significantly to the model for PW in winter only but the final model in each set of regressions was significant for all but WD-CLS in the fall.

Diagnostic Accuracy: ACCESS-W. First, students scoring in the *Beginning* range or below on the ACCESS-W were identified as *at-risk* and coded as '1' in SPSS and those scoring above the *Beginning* range were identified as *no-risk* and coded as '0'. Students not completing the ACCESS-W were not given a code and were removed from analyses. Table 39 provides the numbers of students *at-risk* per grade level according to this process. The number of students that were *at-risk* was very restricted in 3rd grade (17%) and represented the majority in 1st grade (77%). Thus, risk was more readily predicted by grade level than WD in the 1st and 3rd grade. However, further analysis was run for 2nd grade.

Table 39

Numbers of Participants "at-risk" on ACCESS-W

Grade	Total Sample	# at-risk	# no risk	% at-risk
1	21	16	5	76%
2	25	8	17	32%
3	23	4	19	17%

Each predictor and the various combinations (WD-CLS and SAEBRS-AB; PW-WSC and SAEBRS-AB) were entered into logistic regression equations to obtain

predicted probabilities as a first step in determining whether or not SAEBRS-AB added significantly to WD-CLS and PW-WSC, respectively, in the prediction of risk as well as to obtain predicted probabilities for the combined predictors. Diagnostic accuracy was explored for the fall time-point for 2nd grade only with ACCESS-W. Results from the logistic regressions are available in Table 40.

Table 40

Logistic Equation Results for Fall Predictors to ACCESS-W

Model	Screen Measure	*β*	*SE*	Wald	*p*
WD-Fall					
	CLS	-.034	.020	2.765	.096
	AB	.159	.195	.667	.414
	Intercept	-.395	2.394	.027	.869
PW-Fall					
	WSC	-.073	.056	1.734	.188
	AB	.008	.142	.003	.956
	Intercept	1.120	1.981	.320	.572

Note: AB = academic behavior subscale of SAEBRS

In the combined models, none of the predictors were significant in predicting the outcome of risk on the ACCESS-W. Table 41 displays results of the classification accuracy comparisons. AUC is provided as an overall index of accuracy and specificity is provided with cut-points approximating .90, .80, and .70 sensitivity without going below the respective cut-points. Combined predictors were explored using predictive probabilities of combined measures attained via logistic regression. The table also

provides TP, FN, TN, FP, and resulting classification accuracy as a percentage of total sample correctly identified.

Table 41

ROC Curve Analysis with Predictors and ACCESS-W

Measure	*N*	*Sensitivity*	*Specificity*	*Cut Score*	*AUC*	*CI 95%*	*TP*	*FP*	*TN*	*FN*	*Classification Accuracy*
Fall-WD-CLS	22	1.00	.286	131.75	.759*	.546-.972	8	10	4	0	55%
		.875	.50	107.75			7	7	7	1	64%
		.75	.786	66.75			6	3	11	2	77%
Fall-AB	22	.938	.20	17.5	.688	.446-.929	8	11	3	0	47%
		.813	.20	16.5			7	11	3	1	42%
		.75	.60	14.5			6	6	8	2	65%
Fall Combined	22	1.00	.286	.129	.786*	.579-.993	8	10	4	0	55%
		.875	.50	.227			7	7	7	1	64%
		.750	.857	.428			6	2	12	2	82%
Fall-PW-WSC	22	1.00	.40	30.25	.679	.456-.903	8	8	6	0	62%
		.875	.467	29.75			7	7	7	1	62%
		.75	.467	28.25			6	7	7	2	57%
Fall Combined	22	1.00	.467	.274	.675	.452-.898	8	7	7	0	66%
		.875	.467	.278			7	7	7	1	62%
		.75	.467	.296			6	7	7	2	57%

Note: * = $p < .05$; TP = true positives, TN = true negatives, FP = false positives, FN = false negatives

As can be seen in Table 41, the only significant AUC statistics were for WD-CLS and the WD-CLS + SAEBRS-AB predictors. No AUC met the .85 threshold and only

two were above the .70 threshold indicated as 'okay'. Cut-points meeting the .70 sensitivity and specificity thresholds set by Keller-Margulis and colleagues (2016) were with WD-CLS alone (cut-point = 66.75) and the combined WD-CLS + SAEBRS-AB (predicted probability = .428). In comparing WD-CLS alone to the combined WD-CLS + SAEBRS-AB predictors, both had the same sensitivity (.75) but the combined measure had a higher specificity (.857 vs. .786) and better overall classification accuracy (82% vs. 77%). In terms of real implications, this meant that the combined measure resulted in one less FP and one more TN. PW-WSC and the combined PW-WSC + SAEBRS-AB failed to meet the .70 AUC threshold. In summary, WD and the combined WD measure significantly predicted risk on the ACCESS-W for 2nd grade and cut-points meeting the .70 criteria across sensitivity and specificity were met. Differences between WD and the combined WD measure were relatively negligible. No AUC achieved the .85 or better criteria.

Diagnostic Accuracy: MAP-ELA. Only 3rd grade participants completed the MAP-ELA in the spring. For 3rd grade ($N = 19$), students scoring between 230 to 415 are considered *Below Basic* ($N = 10$), from 416-446 as *Basic* ($N = 3$), from 447 to 501 as *Proficient* ($N = 6$), and 502-730 as *Advanced* ($N = 0$). Students scoring in the *Below Basic* range were coded as '1' for *at-risk* and those scoring at *Basic* or better were coded as '0' for *no risk*. Each predictor and the various combinations (WD-CLS and SAEBRS-AB; PW-WSC and SAEBRS-AB) were entered into logistic regression equations to obtain predicted probabilities and as a first step in determining whether or not SAEBRS-AB added significantly to WD-CLS and PW-WSC, respectively, in the prediction of risk. Diagnostic accuracy was explored for the fall and winter time-points only. Results from

the logistic regressions are available in Table 42. In the combined models, none of the predictors were significant in predicting the outcome. Table 43 displays the classification accuracy results following the same model as with the ACCESS-W.

Table 42

Logistic Regression of Combined Predictors to MAP-ELA

Model	Screen Measure	*β*	*SE*	Wald	*p*
WD-Fall	CLS	-.062	.039	2.499	.114
	AB	-.355	.225	2.494	.114
	Intercept	11.406	5.972	3.648	.056
WD-Winter					
	WSC	-.023	.025	.835	.361
	AB	-.569	.334	2.908	.088
	Intercept	10.081	4.628	4.744	.029
PW-Fall					
	WSC	-.135	.087	2.431	.119
	AB	-.419	.225	3.460	.063
	Intercept	8.752	4.257	4.227	.040
PW-Winter					
	WSC	-.639	.497	1.652	.199
	AB	-2.357	1.646	2.050	.152
	Intercept	47.380	34.039	1.937	.164

Table 43

ROC Curve Analysis with Predictors and MAP-ELA

Measure	*N*	*Sensitivity*	*Specificity*	*Cut Score*	*AUC*	*CI 95%*	*TP*	*FP*	*TN*	*FN*	*Classification Accuracy*
Fall-WD-CLS	19	.90	.667	117.75	.811*	.613-1.00	9	3	6	1	79%
		.80	.667	114.75			8	3	6	2	74%
		.70	.778	108.5			7	2	7	3	74%
Fall-AB	19	.90	.667	13.5	.811*	.591-1.00	9	3	6	1	79%
		.80	.778	12.5			8	2	7	2	79%
		.50	.778	11.5			5	2	7	5	63%

Fall WD-Combined	19	.90	.778	.452	.90**	.743-1.00	9	2	7	1	84%
		.80	.889	.509			8	1	8	2	84%
		.70	.889	.561			7	1	8	3	79%
Win-WD-CLS	19	.90	.444	150	.80*	.597-1.00	9	5	4	1	68%
		.80	.444	144.25			8	5	4	2	63%
		.70	.778	131			7	2	7	3	74%
Win-AB	19	1.00	.667	14.5	.878**	.721-1.00	10	3	6	0	84%
		.60	.889	10.5			6	1	8	4	74%
Win WD-Combined	19	.90	.667	.439	.90**	.76-1.00	9	3	6	1	79%
		.80	.667	.487			8	3	6	2	74%
		.70	1.00	.799			7	0	9	3	84%
Fall-PW-WSC	19	.90	.667	30.5	.739	.489-.989	9	3	6	1	79%
		.80	.667	29			8	3	6	2	74%
		.70	.667	25.25			7	3	6	3	68%
Fall PW-Combined	19	.90	.778	.448	.867**	.681-1.00	9	2	7	1	84%
		.80	.778	.533			8	2	7	2	79%
		.70	.778	.617			7	2	7	3	74%
Win-PW-WSC	19	.90	.778	30.75	.783*	.54-1.00	9	2	7	1	84%
		.80	.778	28.75			8	2	7	2	79%
		.70	.778	26.75			7	2	7	3	74%
Win PW-Combined	19	.90	.889	.585	.978***	.922-1.00	9	1	8	1	89%
		.80	1.00	.814			8	0	9	2	89%
		.70	1.00	.881			7	0	9	3	84%

Note: TP = true positives, TN = true negatives, FP = false positives, FN = false negatives, * = $p < .05$; ** = $p < .01$; *** = $p < .001$

As can be seen in Table 43, the AUC was significant for SAEBRS-AB alone in fall and winter, WD-CLS in fall and winter, and PW-WSC in winter. The only AUC that was not significant was for PW-WSC in fall. AUC values ranged from .739 - .878 for the single predictors and .867 - .978 for the combined measures. All combined measures for both fall and winter met the .85 AUC threshold, as did winter SAEBRS-AB. At the preferred sensitivity of .90, only the winter administration of PW-WSC met the .70 criteria for specificity across the individual predictors. However, the combined PW-WSC + SAEBRS-AB met .70 specificity and .90 sensitivity in fall and winter as well as in the fall for the combined WD-CLS + SAEBRS-AB. The combined measures demonstrated the highest overall AUC across time-points (.867 - .978), followed by the SAEBRS-AB (.811 - .878), WD-CLS (.80 - .811), and then PW-WSC (.739 - .783). In summary, the combined measures consistently outperformed any of the single predictors with the combined WD-CLS + SAEBRS-AB performing best in fall and PW-WSC + SAEBRS-AB performing best in winter. However, only the combined PW-WSC + SAEBRS-AB met the .90 sensitivity and .70 specificity criteria across time-points, suggesting that the combined PW-WSC + SAEBRS-AB may be useful for decision-making as it relates to MAP-ELA performance.

CHAPTER 5

DISCUSSION

Overview

The purpose of this study was to examine the technical adequacy of two forms of CBM-W, WD and PW, for ELs in the 1st-3rd grades as well as the utility of combining CBM-W with a measure of academic behavior for the purpose of universal screening. Specifically, the research questions were (1) what is the technical adequacy of WD and PW as indicators of general writing performance for ELs in the 1st-3rd grades (including reliability, validity, sensitivity to growth, and as compared to the general population) and (2) how does the inclusion of a measure of motivated academic behavior in writing impact the predictive validity and diagnostic accuracy of WD and PW for ELs? Discussion by research question will be reported first, followed by limitations, then implications for practice, and ending with future research.

Technical Adequacy of CBM-W

CBM are used for two primary purposes, screening to identify at-risk students and then monitoring a student's responsiveness to instruction across time (Deno, 1985; Deno et al., 2009). To serve these purposes, it is essential that the different forms of CBM are reliable in that they are consistently measuring the same construct across forms and time, and that they are valid in that they are measuring a construct that is highly predictive of performance within a certain domain (e.g., writing). Alternate form reliability was used in this study to establish reliability across forms, time-points, and grades. The results indicated that all production and accurate production metrics for both WD and PW, with the exception of PW-CWS, were reliable across all three grades and time-points for the

ELs in this study ($r > .70$). PW-CWS was reliable across grades and time-points with the exception of the fall for 1st grade ELs ($r = .60$). However, metrics including incorrect sequences had unstable reliability ($r = .60$-$.98$). Alternate form reliabilities were more consistent for the ELs in this study ($r = .60$-97) than reported by McMaster et al. (2014) with the general population ($r = .55$-$.89$). Thus, metrics incorporating accuracy and simple production are the most reliable across forms, grades, and time-points for young ELs in this study. Both forms of CBM-W appear to be more reliable for young ELs than the general population. Reliability, then, does not appear to be a concern when generalizing CBM-W performance of the general population to that of young ELs in this sample. However, PW-CWS may not be the most reliable metric for 1st grade ELs in the fall.

Validity was explored in several ways, including divergent validity, concurrent and predictive convergent validity, and diagnostic accuracy. Ideally, CBM is strongly related to and predictive of a variety of criterion measures within a specific domain (Wayman, Wallace, Wiley, Ticha, & Espin, 2007). For example, studies have established the validity of CBM-R across various standardized assessments of reading comprehension, state assessments of reading comprehension, and more informal teacher evaluations of student reading comprehension (see Wayman et al., 2007). Criterion measures used to establish CBM should have both social and content validity (Wayman et al., 2007). The criterion measures used in this study (ACCESS-W and MAP-ELA) both have social validity as well as varying levels of established criterion validity.

For divergent validity, correlations were consistently stronger with the ACCESS-LC than the ACCESS-ORC for CBM-W metrics that included accuracy (i.e., WSC, CLS,

CWS) while production only metrics (i.e., WW) occasionally correlated more strongly with ACCESS-ORC than ACCESS-LC. This implies that production only metrics may not be able to consistently discriminate between oral English proficiency and writing performance. In other words, metrics incorporating accuracy are better indicators of academic performance while production only metrics are more influenced by oral English proficiency. Thus, metrics incorporating accuracy may be better predictors of future academic performance of ELs across oral English proficiency levels. However, validity results related to the MAP revealed somewhat different patterns than those for the ACCESS-W. Namely, there were no consistent patterns regarding one metric outperforming the other within each form of CBM-W. All PW metrics correlated more strongly with the MAP-ELA than the MAP-MA across time-points while WD did so for the fall and winter time-points only. WD metrics in spring correlated more strongly with MAP-MA than MAP-ELA. This implies that PW is better at discriminating between Mathematics and writing performance for ELs in the 3rd grade, at least at the spring time-point according to the MAP. However, it is worth reiterating that the MAP-ELA incorporated both writing and reading comprehension and did not separate the two constructs in its score reporting.

For concurrent and predictive validity, writing has proven to be a difficult construct to measure, with validity coefficients between standardized assessments in writing reported as .30 to .50, as compared to .85 in reading (Romig, et al., 2017). Thus, CBM-W can be expected to have lower validity coefficients than CBM-R (r = .70 to .80; Wayman et al., 2007). WD-WSC and CLS correlated more strongly with the ACCESS-W (r = .56 - .81) than PW-WSC or CWS (r = .15 - .62) across grades and time-points.

However, PW-WSC and CWS correlated more strongly with the MAP-ELA (r = .44 - .67) than WD-WSC or CLS did (r = .35 - .44). Thus, the respective validity of each form of CBM-W changes in relation to the outcome measure used but validity coefficients should be considered promising to strong given the construct being measured. Between the ACCESS-W and MAP-ELA, the ACCESS-W is designed to specifically assess writing as a separate construct while the MAP-ELA conflated both writing and reading comprehension and EL performance on the ACCESS-W has greater impact upon the EL in terms of access to and type of ESOL services provided (e.g., social validity), thus results related to the ACCESS-W have more validity than those with the MAP-ELA. However, this also suggests that PW may be more predictive of reading comprehension skills than WD. Considering that PW is designed to assess both transcription and text generation skills (i.e., expressive vocabulary and basic grammatical knowledge), then higher coefficients with reading comprehension for PW makes sense because WD is designed to assess transcription level skills only, which relate more specifically to decoding skills and not necessarily vocabulary knowledge. In accordance, previous studies have consistently underscored the significance of vocabulary knowledge for EL literacy and suggest PW may have predictive properties that go beyond writing performance alone (August et al., 2014; Genesse et al., 2005).

WD-WSC and CLS correlated within a similar range for the general population using the same CBM-Ws (r = .41-.83; McMaster et al., 2014) as it did for ELs in this study (r = .52-.74), although the studies employed different criterion measures. For the general population, PW-WSC had a range of r = -.04 to .58 with the lowest correlations being in 1st grade (Allen et al., n.d.) while PW-WSC ranged from r = .19 to .66 with

lowest correlations in the 1st grade for ELs in this study. In their meta-analysis, Romig and colleagues (2017) found that WSC had an overall correlation of .44 while CWS had an overall correlation of .51. This indicates that both forms of CBM-W had similar patterns of correlations across various outcomes measures for both ELs and the general population for this study. This suggests that both WD and PW are just as valid for use with ELs as with the general population, especially when using metrics that incorporate accuracy.

Another indicator of technical adequacy is diagnostic accuracy (Johnson et al., 2007). For purposes of screening, it is essential that CBM accurately identify students that are in need of early intervention and ROC curve analysis is a commonly applied method for establishing diagnostic accuracy (Johnson et al., 2007). Lack of variance across the categories of 'no risk' and 'at risk' according to the ACCESS-W for 1st and 3rd grade precluded a more detailed analysis, but analyses were conducted for 2nd grade with the ACCESS-W and 3rd grade for the MAP-ELA. The ACCESS-W was administered in winter for 2nd grade. WD-CLS (AUC = .759) outperformed PW-WSC (AUC = .679) in predicting risk for 2nd grade ELs on the ACCESS-W. The MAP-ELA was administered in spring to 3rd grade ELs. Again, WD-CLS (AUC = .811) outperformed PW-WSC (AUC = .739). This suggests that WD-CLS is better than PW-WSC at identifying risk for the grades analyzed and criterion measures employed in this study. Thus, WD-CLS has emerging evidence that it may be used for screening purposes to identify young ELs that are in need of early intervention in writing. This is critical because of current trends in under-identifying young ELs for early intervention combined with evidence that early intervention can promote academic success for young ELs (NAESM, 2017).

Beyond the validity of a static score, sensitivity to growth is a key component of CBM because it is used to monitor a student's rate of learning in response to instruction across time (Deno et al., 2009). Sensitivity to growth has implications for CBM-W's utility for progress monitoring. Sensitivity to growth was analyzed across all metrics descriptively and indicated that both forms of CBM-W and each metric increased across time-points within grade as well as across grades with the exception of 2nd grade ELs scoring slightly higher than 3rd grade ELs in the fall for PW-WSC and CWS. This means that both forms of CBM-W demonstrate some sensitivity to growth across time.

Only WD-WSC, WD-CLS, PW-WSC, and PW-CWS were examined statistically for sensitivity to growth. The justification for examining these metrics was that metrics incorporating accuracy consistently had better validity, both convergent and divergent, and appeared to be less influenced by oral English proficiency. Statistical analyses indicated that WD-WSC was sensitive to growth across each time-point within each grade while WD-CLS, PW-WSC, and PW-CWS were statistically significantly different across each time-point for 1st and 3rd grade ELs. Only 2nd grade ELs had mixed patterns of significant growth according to WD-CLS, PW-WSC, and PW-CWS. There was no significant growth from winter to spring for WD-CLS and fall to winter for either PW-WSC or PW-CWS.

Lack of significant growth across time-points for 2nd grade ELs could be a matter of sample characteristics and small sample size or could indicate key times for ELs in relation to writing growth in response to general instruction in writing. For example, ELs may be slowing down their rate of production to focus more upon accuracy of their production during the 2nd grade. Furthermore, both WD and PW discriminated between

1st grade ELs and 2nd grade ELs as well as between 1st grade and 3rd grade ELs across metrics. However, none of the metrics statistically significantly discriminated between 2nd and 3rd grade ELs. Again, differences between grades could possibly be a function of the sample size and sample characteristics or indicate a key time during which ELs change their focus in writing from production to accuracy. Taken together, preliminary evidence indicates that both WD and PW are sensitive to growth over time for young ELs, especially since growth happened within the context of business as usual and research has indicated that writing instruction is often neglected in the early grades (i.e., not much growth would be expected because writing is not a focus of instruction; National Commission on Writing, 2003).

Across all of the analyses for reliability and validity, WD-CLS is the most appropriate form and metric for 1st grade ELs, due to lower reliability by PW and floor effects for WD-WSC, and 2nd grade ELs up to at least the winter time-point. Results according to sensitivity to growth suggest that WD-WSC may be a more consistent indicator of growth across time-points for 2nd grade ELs. Correlational analyses and diagnostic accuracy for 2nd and 3rd grade ELs is a bit mixed. WD outperformed PW across grades and time-points with the ACCESS-W and had better diagnostic accuracy with the MAP-ELA despite weaker correlations than PW for 3rd grade ELs. Therefore, WD appears to be the most robust form across grades and time-points, at least in so far as writing performance on the criterion measures used in this study. However, PW still performed moderately well for 2nd and 3rd grade ELs and remains a viable alternative for screening and progress monitoring for students struggling with text-generation skills as opposed to transcription skills.

EL performance across time points was compared descriptively to the general population to determine if seasonal benchmarks and rates of growth drawn from the general population may be applicable to ELs, a common concern in the current literature (Abedi & Garanda, 2006; Burr et al., 2015). Average EL performance consistently fell within a standard deviation of average performance by the general population with the exception of 3rd grade ELs, for which WD-WSC at each time-point and WD-CLS in fall and winter was beyond a standard deviation below the general population. Patterns of EL performance falling further behind that of their non-EL peers in 3rd grade could indicate a critical juncture during which benchmarks and rates of growth are no longer applicable to beginning and intermediate oral English proficiency ELs or it could be an anomaly related to the small sample size. Also, 2nd grade ELs actually attempted more words and spelled more words correctly than the general population in the fall. This may be a function of ESL instruction at this time, wherein many ELs begin learning and mastering basic sentence frames (i.e., "I like.." or "I love..") that they may over-apply on PW in conjunction with circumlocution, which is using several different words that approximate a word not in the individual's expressive vocabulary, that can result in an increased number of words written (Arias, 2008). However, this descriptive analysis indicates that benchmarks and rates of growth drawn from the general population may be cautiously applied with young ELs for purposes of identifying risk, creating long-term goals, and monitoring progress across time.

Finally, as it relates to prior studies examining CBM-W with ELs, the results related to reliability and validity are in line with those conducted with older ELs (Campbell, 2010; Campbell et al., 2013; & Espin et al., 2008) in that validity is similar to,

if not stronger, for ELs as compared to non-ELs. However, the findings here are in direct conflict with the only other study with young ELs by Keller-Margulis and colleagues (2016) because they found validity to be stronger for non-ELs than ELs. The primary difference between this study and that conducted by Keller-Margulis et al. (2016) is that they employed CBM-SP and this study used CBM-WD and PW. Therefore, it seems that either WD or PW are more appropriate measures for screening and progress monitoring young ELs in writing.

Utility of Combining CBM-W and Academic Behavior

Integrating measures of motivated academic behavior and CBM-W may improve the predictive validity and diagnostic accuracy of CBM-W alone because CBM-W does not account for an EL's prior academic experiences or opportunities to learn. Again, ELs come to the US education system with diverse prior academic experiences in terms of their time receiving formal education in both English and their L1 (Artiles et al., 2005). A measure of motivated academic behavior may not take into account these prior experiences directly, but it does provide information related to an individual student's likelihood of taking advantage of future opportunities to learn via on-task and engaged behavior (DiPerna & Elliott, 2002). Furthermore, current academic behavior is energized via motivation, which is developed based upon an individual's prior learning experiences and values (e.g., motivation; Boscolo & Gelati, 2006). The utility of combining CBM-W with the SAEBRS-AB, a measure of motivated academic behavior, was explored in two ways. First, hierarchical regression was used to examine how much additional variance was explained in criterion measure performance by SAEBRS-AB when either WD or PW was entered into the model first. Second, a series of ROC curve analyses were run for

each predictor alone as well as for a combined score that incorporated both SAEBRS-AB and WD or PW for predicting risk on each criterion measure. The various methods of prediction were compared to each other according to their respective AUC, sensitivity, specificity, and overall diagnostic accuracy.

The addition of the SAEBRS-AB did not significantly improve the predictive performance of WD-CLS when accounting for grade to the ACCESS-W, but, in contrast to WD, the addition of SAEBRS-AB added significantly to the predictive performance of PW-WSC to the ACCESS-W. Oral English proficiency is a significant predictor of ACCESS-W performance for both WD and PW and should be accounted for in future studies. However, results were somewhat different when MAP-ELA was used as the criterion measure. The final model in fall for WD was not significant ($p > .05$) but the addition of PW-WSC in fall was significant ($p < .05$). Furthermore, the addition of the SAEBRS-AB to the final model for winter PW-WSC was significant ($p < .05$) and the final model for fall PW-WSC and SAEBRS-AB was also significant ($p < .05$), although the increase in R^2 when the SAEBRS-AB was added was not significant ($p > .05$). This suggests that the addition of SAEBRS-AB has more utility for PW than WD but the best form of CBM-W changes as a function of the criterion measure employed. Future studies should continue exploring the differing validity across a variety of criterion measures in order to identify the most consistent form of CBM-W.

Existing literature consistently highlights oral English proficiency as a key variable in EL literacy performance (Abedi & Garanda, 2006; Genesse et al., 2005) and results related to the ACCESS-W support those claims, but results related to the MAP-ELA do not. These mixed finding may be a product of the small sample size or could be a

function of the criterion measures. The ACCESS-W was created as a subtest within a larger test specifically created to assess an EL's English proficiency so it stands to reason that the two sub-tests (ACCESS-ORC and ACCESS-W) are highly related, as is substantiated by technical reports for the ACCESS (Yanosky et al., 2013). However, this does not necessarily mean the two constructs are highly related, only that the two sub-tests are highly related. In contrast, the MAP-ELA was not created with oral English proficiency in mind nor was oral English proficiency a sub-test for the MAP. If oral English proficiency does influence writing performance and it is adequately measured by the ACCESS-ORC, then oral English proficiency should be a significant contributor to EL writing performance across criterion measures. The final model of PW-WSC plus SAEBRS-AB in fall was a significant predictor of criterion measure performance, either ACCESS-W or MAP-ELA, across time-points analyzed without the need of including oral English proficiency. WD-CLS plus the SAEBRS-AB significantly predicted performance in fall for ACCESS-W but not MAP-ELA. Again, this raises more questions regarding the significance of initial oral English proficiency, at least for ELs in the 1st through 3rd grade. Thus, findings do suggest that combining the SAEBRS-AB with either WD-CLS or PW-WSC is a more robust predictor of writing performance than oral English proficiency, although oral English proficiency should still be accounted for in analyses until the relationship is better understood. Furthermore, SAEBRS-AB explained more variance when combined with PW-WSC than with WD-CLS across criterion measures. This could be because WD-CLS is simply a more valid measure of writing performance than PW-WSC.

The utility of combining CBM-W and SAEBRS-AB was also explored via ROC curve analyses, which is a non-parametric procedure. ROC curve analyses were conducted for 2nd graders using the ACCESS-W and 3rd graders with the MAP-ELA. For 2nd grade with the ACCESS-W, only WD-CLS and the WD-CLS combined with the SAEBRS-AB had an AUC above .70 (.76 and .79, respectively), but neither met the .85 cut-point for convincing evidence. WD-CLS alone performed very equitably to the combined WD-CLS and SAEBRS-AB measure but, when examining cut score performance with .70 thresholds for both sensitivity and specificity, the combined cut-score had better classification accuracy (82% vs. 77%) that resulted in one less false positive without sacrificing true positives. Thus, there is value added by combining the measures when predicting risk on the ACCESS-W.

For the MAP-ELA, all combined measures had a higher AUC than WD-CLS, PW-WSC, or SAEBRS-AB alone. The AUC was excellent (greater than or equal to .90) for fall WD-CLS and SAEBRS-AB combined, winter WD-CLS and SAEBRS-AB combined, and winter PW-WSC and SAEBRS-AB combined. Furthermore, cut-scores with a .90 sensitivity and .70 specificity minimum were found using the combined measures for both WD and PW in both fall and winter. According to Jenkins et al. (2007), this meets minimum requirements for decision-making and supports the utility of combining CBM-W with SAEBRS-AB for identifying risk for 3rd grade ELs, at least as it pertains to performance on the MAP-ELA. Again, the ability to accurately identify risk and provide early intervention supports to young ELs is important for promoting academic success of ELs (NAESM, 2017) and results suggest combining the SAEBRS-AB with CBM-W lead to more accurate identification of risk.

In summary, ELs represent a diverse and high-risk sub-population of students that have been traditionally under-identified for early intervention services (NAESM, 2017). Furthermore, NAESM (2017) stress the use of CBM for early identification and progress monitoring ELs in response to instruction. This study lends support to WD and PW as valid tools for screening and progress monitoring ELs in the 1st through 3rd grades. Moreover, benchmarks and rates of growth drawn from the general population may be tentatively applied to young ELs, at least those with beginning to intermediate oral English proficiencies. Furthermore, the sample included native speakers of 11 different languages, therefore this study provides support for both WD and PW as a reliable and valid measures across native languages. The integration of the SAEBRS-AB and CBM-W improves diagnostic accuracy and may also have value in informing initial choices in intervention. For example, students that screen as 'at-risk' according to both the SAEBRS-AB and CBM-W may benefit from interventions that explicitly target both motivated academic behaviors as well as academic skills while those that only score as 'at-risk' according to CBM-W may not need the same emphasis upon motivated academic behavior within their intervention. This integrated model of screening may help schools more effectively allocate resources to individual students while maximizing student outcomes.

Limitations

Several limitations impact the generalizability of this study and serve to caution the interpretation of the findings. The first limitation is the small sample size, which limited the analysis of the impact of oral English proficiency upon the reliability and validity of CBM-W. Furthermore, the majority of participants were in the beginning to

intermediate range of oral English proficiency and there were no 2015/2016 ACCESS results for nearly one third of the sample. Moreover, for those for whom 2015/2016 ACCESS results were made available, this represented their English proficiency in winter of 2015 which is certainly subject to change over the course of the spring school semester, summer, and early fall prior to fall benchmarking in November. Thus, it is impossible to determine how initial oral English proficiency influenced CBM reliability and validity in the fall of this study or whether or not these findings could be generalized to ELs with advanced levels of oral English proficiency. Another limitation related to the sample was the lack of socio-economic data. The school district would not allow the collection of free/reduced lunch status, which precluded the examination of socio-economic status in the analyses. Again, ELs represent an extremely heterogeneous group of students and, although many L1s were represented within this study, any generalization to all ELs should be done with caution. Furthermore, there was no data regarding the participants' proficiency in their L1.

Beyond the limitations related to the sample size and available demographic data, there were also limitations regarding the criterion measures. The ACCESS was administered on-line for the first time in 2016/2017, which was the year in which this study was conducted. It is unclear how the on-line administration may have impacted student performance. Furthermore, although the ACCESS is a socially valid assessment for ELs, it is not an assessment used to identify risk or eligibility for special. Thus, the cut-score for risk was selected by the author and may have limited utility. The MAP-ELA is an assessment of both reading and writing and does not separate the two constructs. Finally, a lack of variance in rates of risk for 1st and 3rd grade with the ACCESS-W and

no data for 1st or 2nd grade with the MAP-ELA precluded a more detailed analysis of diagnostic accuracy across grades or with multiple criterion measures.

Implications for Practice

Although the results presented in this study should be considered exploratory in nature and it is not advisable that CBM-W data be used for placement decisions at this time, teachers may still use WD and PW data as a method of informing in-class groupings, targeting writing instruction, creating goals, and monitoring student progress. In general, teachers may use published benchmarks for the general population as guidelines in terms of general performance for goal creation and targeting instruction with young ELs but they should carefully consider the individual student's oral English proficiency and prior formal education in both English and their L1, at a minimum, prior to referral for special education or placement decisions within a MTSS. Essentially, more research is needed before CBM-W data should be used for placement decisions, although it may be included as part of the data and evidence for such decisions. Most importantly, practitioners can use both forms of CBM-W to help them identify young ELs in need of regular progress-monitoring and as potentially needing increased instructional time in writing within the scope of their general ESOL instruction at the beginning of the school year, rather than waiting for achievement gaps to grow across the year.

Future Research

This study raises several questions that should be addressed in future research. First, the technical adequacy of both WD and PW need further exploration using a variety of criterion measures that should include standardized assessments as well as teacher evaluations. As noted in the limitations, both criterion measures used within this study

are not beyond question and assessments for universal screening and progress monitoring should remain robust across a variety of criterion measures. Beyond replicating this study with predominately beginning and intermediate proficiency ELs, future studies should explore the adequacy of WD and PW with advanced proficiency and monitored ELs. Other variables that warrant future study are the student's time receiving formal English instruction, L1 fluency, and socio-economic status. Also, studies should examine the adequacy of CBM-W within the context of bi-lingual models of instruction since research has shown learning trajectories change as a function of an EL's model of instruction (Slavin et al., 2005) as well as create and validate CBM-Ws in different languages.

Results from this study regarding the influence of oral English proficiency upon the validity of CBM-W are mixed and future studies should explicitly explore the importance initial oral English proficiency plays in the prediction of future performance by CBM-W. It is highly likely that initial oral English proficiency plays an increasingly critical role in the identification of risk and predictive capabilities of CBM-W as the student ages. In other words, using benchmarks drawn from the general population may be acceptable for ELs in the 1st or 2nd grade but may result in over-identification of risk and inappropriate placement decisions in the middle to upper-elementary grades. Essentially, the discrepancy between the EL student and the general population, as well as the import of that discrepancy, is likely different for a 1st grade EL with beginning proficiency as compared to a 3rd grade EL with beginning proficiency. Furthermore, studies should explore the utility of common ELP assessments, such as the ACCESS, in order to determine whether or not they provide results that are meaningful and helpful to the students that take it as well as the teachers and schools that administer it.

Although this study provides some evidence related to WD and PW's sensitivity to growth, future studies should explore slope validity in order to establish the minimum number of data-points needed to project academic growth; specific factors related to ELs that may influence slope and rates of improvement, as research with reading CBM indicates differential growth patterns for ELs (Keller-Margulis et al., 2012); and establish decision rules that teachers can use to meaningfully modify instruction and/or make placement decisions within a MTSS framework. Future studies should explicitly examine whether or not WD and/or PW, when used within a DBI framework, actually result in improved academic performance by young ELs and increase their access to educational opportunities and resources. If the use of CBM-W does not result in improved academic outcomes and instructional practices by teachers of ELs then it is nothing more than just another way to document the already well-documented achievement gap. Finally, research should continue to explore the utility of combining CBM and behavioral data for screening purposes as well as for use in identifying interventions corresponding to student performance across measures. Beyond more efficient early identification of risk, it is hypothesized here that the integration of measures could be used to generate student profiles that can then be matched to interventions that are better aligned with the individual student's specific needs.

REFERENCES

Abbott, R. D., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology*, *102*(2), 281-298.

Abedi, J., & Gándara, P. (2006). Performance of English Language Learners as a Subgroup in Large-Scale Assessment: Interaction of Research and Policy. *Educational Measurement: Issues and Practice*, *25*(4), 36-46.

Abu-Hamour, B. (2013). Arabic spelling and curriculum based measurement. The Australian Educational and Developmental Psychologist, 30(02), 140-156.

Abu-Hamour, B. (2014). A pilot study for standardizing curriculum-based measurement oral reading fluency (CBM ORF) in Arabic. Journal of the International Association of Special Education, 15(1) 16-26.

Abu-Hamour, B., Al-Hmouz, H., & Kenana, M. (2013). The effect of short vowelization on curriculum-based measurement of reading fluency and comprehension in Arabic. Australian Journal of Learning Difficulties, 18(2), 181-197.

Artiles, A. J., Rueda, R., Salazar, J. J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children*, *71*(3), 283-300.

Aud, S., Hussar, W., Johnson, F., Kena, G., Roth, E., Manning, E., ... & Zhang, J. (2012). The Condition of Education 2012. NCES 2012-045. *National Center for Education Statistics*.

August, D., McCardle, P., & Shanahan, T. (2014). Developing literacy in English language learners: Findings from a review of the experimental research. *School Psychology Review*, *43*(4), 490-498.

August, D. & Shanahan, T. (Eds.). (2006). Developing literacy in second language learners: A report of the National Literacy Panel on Language Minority Children and Youth, executive summary, Center for Applied Linguistics. Mahwah, NJ: Lawrence Erlbaum. http://www.cal.org/projects/archive/nlreports/Executive_Summary.pdf.

August, D., Shanahan, T., & Escamilla, K. (2009). English language learners: Developing literacy in second-language learners—Report of the National Literacy Panel on Language-Minority Children and Youth. *Journal of Literacy Research*, *41*(4), 432-452.

Babayiğit, S. (2014). Contributions of word-level and verbal skills to written expression: comparison of learners who speak English as a first (L1) and second language (L2). *Reading and Writing*, *27*(7), 1207-1229.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York, NY: Freeman.

Bazerman, C. (2016). What do sociocultural studies in writing tell us about learning to write? In C.A., McArthrur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research 2nd* edition (pp. 11-23). New York, NY: The Guilford Press.

Berninger, V. W., & Abbott, R. D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language

systems in grades 1, 3, 5, and 7. *Journal of educational psychology*, *102*(3), 635-651.

Berninger, V. W., & Amtmann, D. (2003). Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice. In H. Swanson, K. Harris, and S. Graham (Eds.) *Handbook of Learning Disabilities* (pp. 323-344). New York: The Guilford Press.

Berninger, V. W., Rutberg, J. E., Abbott, R. D., Garcia, N., Anderson-Youngstrom, M., Brooks, A., & Fulton, C. (2006). Tier 1 and tier 2 early intervention for handwriting and composing. *Journal of School Psychology*, 44, 3-30.

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., Hawkins, J. M., & Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology*, 94, 291-304.

Betts, J., Bolt, S., Decker, D., Muyskens, P., & Marston, D. (2009). Examining the role of time and language type in reading development for English language learners. *Journal of School Psychology*, 47(3), 143-166.

Boscolo, P. (2014). Two metaphors for writing research and their implications for writing instruction. In B. Arfe, J. Dockrell, & V. Berninger (Eds.), *Writing development in children with hearing loss, dyslexia, or oral language problems: Implications for assessment and instruction.* (pp. 33-44). New York, NY: Oxford University Press.

Boscolo, P. & Gelati, C. (2007). Best practices in promoting motivation for writing. In S. Graham, C.A., MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (pp. 202-221). New York, NY: The Guilford Press.

Boscolo, P. & Hidi, S. (2007). The multiple meanings of motivation to write In G. Rijlaarsdam (Series Ed.) and P. Boscolo and S. Hidi (Volume Eds.), *Studies in Writing, Volume 19*, *Writing and Motivation* (pp. 1-14). Oxford: Elsevier.

Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology*, *105*(1), 25-38.

Bruning, R., & Horn, C. (2000). Developing motivation to write. *Educational psychologist*, *35*(1), 25-37.

Bruning, R.H., & Kauffman, D. F. (2016). Self-efficacy beliefs and motivation in writing development. In MacArthur, C.A., Graham, S., & Fitzgerald, J. (Eds.), *Handbook of writing research 2nd edition.* (pp. 160-173). New York, NY: The Guilford Press.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42-65.

Burr, E., Haas, E., & Ferriere, K. (2015). Identifying and Supporting English Learner Students with Learning Disabilities: Key Issues in the Literature and State Practice. REL 2015-086. *Regional Educational Laboratory West*.

Campbell, H.M. (2006). The technical adequacy of curriculum based measurement in writing with English language learners. (unpublished doctoral dissertation). University of Minnesota, MN.

Campbell, H. M. (2010). The technical adequacy of curriculum-based measurement passage copying with secondary school English language learners. *Reading & Writing Quarterly*, 26(4), 289-307.

Campbell, H., Espin, C. A., & McMaster, K. (2013). The technical adequacy of curriculum-based writing measures with English learners. *Reading and Writing*, 26(3), 431-452.

Capizzi, A. M., & Fuchs, L. S. (2005). Effects of curriculum-based measurement with and without diagnostic feedback on teacher planning. *Remedial and Special Education*, *26*(3), 159-174.

Catts, H. W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of learning disabilities*, *42*(2), 163-176.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing generating text in L1 and L2. *Written Communication*, 18(1), 80-98.

Ritchey, K. D., & Coker Jr, D. L. (2013). An investigation of the validity and utility of two curriculum-based measurement writing tasks. *Reading & Writing Quarterly*, *29*(1), 89-119.

Crawford, J. (2004). No Child Left Behind: Misguided approach to school accountability for English language learners. In *Center on Educational Policy's Forum on Ideas to Improve the NCLB Accountability Provisions for Students with Disabilities and English Language Learners.* Retrieved from: http://www.nabe.org/resources/documents/nclb%20page/nabe_on_nclb.pdf

Clemens, N. H., Shapiro, E. S., & Thoemmes, F. (2011). Improving the efficacy of first grade reading screening: An investigation of word identification fluency with other early literacy indicators. *School Psychology Quarterly*, *26*(3), 231.

Coie, J.D., Watt, N.F., West, S.G., Hawkins, D., Asarnow, J.R., Markman, H.J., Ramey, S.L., Shure, M.B., & Long, B. (1993). The science of prevention. *American Psychologist.* 48, 1013-1022.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., ... & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of educational psychology*, *102*(2), 327.

Connelly, V. (2014). Integrating writing and oral language disorders: Perspectives of a writing researcher. In B. Arfe, J. Dockrell, & V. Berninger (Eds.), *Writing development in children with hearing loss, dyslexia, or oral language problems: Implications for assessment and instruction. (pp. 313-324).* New York, NY: Oxford University Press.

Conoyer, S. J., Foegen, A., & Lembke, E. S. (2016). Early numeracy indicator: Examining predictive utility across years and states. *Remedial and Special Education, 37*(3), 159–171. doi:10.1177/0741932515619758

Cumming, A. (2016). Writing development and instruction for English language learners. In C.A., McArthrur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research 2nd* edition (pp. 364-376). New York, NY: The Guilford Press.

Cumming, A., Kim, T.-Y., & Eouanzoui, K.B. (2007). Motivation for esl writing improvement in pre-university contexts In G. Rijlaarsdam (Series Ed.) and P. Boscolo and S. Hidi (Volume Eds.), *Studies in Writing, Volume 19*, *Writing and Motivation* (pp. 93-111). Oxford: Elsevier.

Cummins, J. (1980). The construct of language proficiency in bilingual education. *Current issues in bilingual education*, 81-103.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy* (Vol. 6). Clevedon: Multilingual Matters.

Danzak, R. L. (2011). The integration of lexical, syntactic, and discourse features in bilingual adolescents' writing: An exploratory approach. *Language, Speech, and Hearing Services in Schools*, *42*(4), 491-505.

Danzak, R.L., & Silliman, E.R., (2014). Writing development of Spanish-English bilingual students with language learning disabilities: New directions in constructing individual profiles. In B. Arfe, J. Dockrell, & V. Berninger (Eds.), *Writing development in children with hearing loss, dyslexia, or oral language problems: Implications for assessment and instruction. (pp. 158-175).* New York, NY: Oxford University Press.

de Jong, E. J. (2011). *Foundations for multilingualism in education: From principles to practice*. Philadelphia, PA: Caslon Publishing.

de Jong, E. J. (2013). Policy discourses and US language in education policies. *Peabody Journal of Education*, *88*(1), 98-111.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional children*, *52*(3), 219-232.

Deno, S. L., & Mirkin, P. K. (1977). Data-based program modification: A manual. Reston VA: Council for Exceptional Children.

Deno, S. L., Mirkin, P. K., & Marston, D. (1980). Relationships among simple measures of written expression and performance on standardized achievement tests (Rep. No. 22). *Minneapolis, Minnesota*.

Deno, S. L., Reschly, A. L., Lembke, E. S., Magnusson, D., Callender, S. A., Windram, H., & Stachel, N. (2009). Developing a school-wide progress-monitoring system. *Psychology in the Schools*, *46*(1), 44-55.

DiPerna, J. C., & Elliott, S. N. (2002). Promoting academic enablers to improve student achievenment: An introdution to the mini-series. *School Psychology Review*, *31*(3), 293-297.

DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2002). A model fo academic enablers and elementary reading/language arts achievement. *School Psychology Review*, *31*(3), 298-312.

Dockrell, J.E., & Arfe, B. (2014). The role of oral language in developing written language skills: Questions for European pedagogy? In B. Arfe, J. Dockrell, & V. Berninger (Eds.), *Writing development in children with hearing loss, dyslexia, or oral language problems: Implications for assessment and instruction. (pp. 325-335).* New York, NY: Oxford University Press.

Dressler, C. & Kamil, M.L. (2006). First- and second-language literacy instruction. In D. August & T. Shanahan (Eds), *Developing literacy in second language learners* (pp 197-238). Mahwah, NJ: Lawrence Erlbaum.

Espin, C., Wallace, T., Campbell, H., Lembke, E. S., Long, J. D., & Ticha, R. (2008). Curriculum-based measurement in writing: Predicting the success of high-school students on state standards tests. *Exceptional Children*, *74*(2), 174-193.

Evans, B. A., & Hornberger, N. H. (2005). No child left behind: Repealing and unpeeling federal language education policy in the United States. *Language Policy*, *4*(1), 87-106.

Ferdman, B. (1990). Literacy and cultural identity. *Harvard Educational Review*, *60*(2), 181-205.

Fitzgerald, J., & Amendum, S. (2007). What is sound writing instruction for multilingual learners. *Best practices in writing instruction*, 289-307.

Fox, J., & Fairbairn, S. (2011). Test Review: ACCESS for ELLs [R]. *Language Testing*, *28*(3), 425-431.

Fry, R. (2010). Hispanics, high school dropouts and the GED. Washington, DC: Pew

Hispanic Center.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, *33*(2), 188-193.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it?. *Reading research quarterly*, *41*(1), 93-99.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children*, *78*(3), 263-279.

Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English language learners in US schools: An overview of research findings. *Journal of Education for Students Placed at Risk*, *10*(4), 363-385.

Gersten, R., Clarke, B., Jordan, N. C., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. *Exceptional Children, 78*(4), 423–445.

Graham, S., Berninger, V., & Fan, W. (2007). The structural relationship between writing attitude and writing achievement in first and third grade students. *Contemporary educational psychology*, *32*(3), 516-536.

Graham, S., & Harris, K. R. (2009). Almost 30 years of writing research: Making sense of it all with The Wrath of Khan. *Learning Disabilities Research & Practice*, *24*(2), 58-68.

Graham, S., Harris, K. R., & Larsen, L. (2001). Prevention and intervention of writing difficulties for students with learning disabilities. *Learning Disabilities Research & Practice*, *16*(2), 74-84.

Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, *81*(4), 710-744.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of educational psychology*, *99*(3), 445-476.

Hampton, D. D., & Lembke, E. S. (2016). Examining the technical adequacy of progress monitoring using early writing curriculum-based measures. *Reading & Writing Quarterly*, *32*(4), 336-352.

Harrison, G. L., Goegan, L. D., Jalbert, R., McManus, K., Sinclair, K., & Spurling, J. (2016). Predictors of spelling and writing skills in first-and second-language learners. *Reading and Writing*, *29*(1), 69-89.

Hresko, W. P., Herron, S. R., & Peak, P. K. (1996). *Test of early written language-2 (TEWL-2)*. Austin, TX: ProEd.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review*, *36*(4), 582.

Kaminitz-Berkooz, I., & Shapiro, E. S. (2005). The applicability of curriculum-based measurement to measure reading in Hebrew. *School Psychology International*, 26(4), 494-519.

Keller-Margulis, M. A., Clemens, N. H., Im, M. H., Kwok, O. M., & Booth, C. (2012). Curriculum-based measurement yearly growth rates: An examination of English language learners and native English speakers. *Learning and Individual Differences*, *22*(6), 799-805.

Keller-Margulis, M., Payan, A., Jaspers, K. E., & Brewton, C. (2016). Validity and Diagnostic Accuracy of Written Expression Curriculum-Based Measurement for Students With Diverse Language Backgrounds. *Reading & Writing Quarterly*, *32*(2), 174-198.

Keller-Margulis, M. A., Payan, A., & Booth, C. (2012). Reading Curriculum-Based Measures in Spanish An Examination of Validity and Diagnostic Accuracy. *Assessment for Effective Intervention*, *37*(4), 212-223.

Kena, G., Hussar W., McFarland J., de Brey C., Musu-Gillette, L., Wang, X., Zhang, J., Rathbun, A., Wilkinson-Flicker, S., Diliberti M., Barmer, A., Bullock Mann, F., and Dunlop Velez, E. (2016). The Condition of Education 2016 (NCES 2016-144). U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved [2016] from http://nces.ed.gov/pubsearch

Kilgus, S. P., Chafouleas, S. M., & Riley-Tillman, T. C. (2013). Development and initial validation of the Social and Academic Behavior Risk Screener for elementary grades. *School Psychology Quarterly*, *28*(3), 210-226.

Kilgus, S. P., Eklund, K., Nathaniel, P., Taylor, C. N., & Sims, W. A. (2016). Psychometric defensibility of the Social, Academic, and Emotional Behavior Risk

Screener (SAEBRS) Teacher Rating Scale and multiple gating procedure within elementary and middle school samples. *Journal of School Psychology*, *58*, 21-39.

Kilgus, S. P., Sims, W. A., Nathaniel, P., & Taylor, C. N. (2016). Technical Adequacy of the Social, Academic, and Emotional Behavior Risk Screener in an Elementary Sample. *Assessment for Effective Intervention*, DOI: 10.1177/1534508415623269.

Kindler, A. L. (2002). Survey of the states' limited English proficient students and available educational programs and services: 2000–2001 summary report. Washington, DC: *National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs*.

Kinloch, V. & Burkhard, T. (2016). Teaching writing in culturally and linguistically diverse classrooms. In C.A., McArthrur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research 2nd* edition (pp. 377-394). New York, NY: The Guilford Press.

Klassen, R. (2002). Writing in early adolescence: A review of the role of self-efficacy beliefs. *Educational psychology review*, *14*(2), 173-203.

Klingner, J. K., Artiles, A. J., & Barletta, L. M. (2006). English Language Learners Who Struggle With Reading Language Acquisition or LD?. *Journal of Learning Disabilities*, *39*(2), 108-128.

LeRoy, M. (Producer), & Fleming, V. (Director). (1939). *The Wizard of Oz* [Motion picture on DVD]. United States: MGM.

Lembke, E., Deno, S. L., & Hall, K. (2003). Identifying an indicator of growth in early writing proficiency for elementary school students. *Assessment for Effective Intervention*, *28*(3-4), 23-35.

Lembke, E. S., McMaster, K. L., & Stecker, P. M. (2010). The prevention science of reading research within a Response-to-Intervention model. *Psychology in the Schools, 47*(1), 22-35. doi:10.1002/pits.20449

Linan-Thompson, S., Cirino, P. T., & Vaughn, S. (2007). Determining English language learners' response to intervention: Questions and some answers. *Learning Disability Quarterly*, *30*(3), 185-195.

Linquanti, R. & Cook, H.G. (2015). *Re-examining reclassification: Guidance from a national working session on policies and practices for exiting students English learner status.* Washington, DC: Council of Chief State School Officers.

Logan, J. A., & Petscher, Y. (2010). School profiles of at-risk student concentration: Differential growth in oral reading fluency. *Journal of School Psychology*, 48(2), 163-186.

Matute-Bianchi, M. E. (1986). Ethnic identities and patterns of school success and failure among Mexican-descent and Japanese-American students in a California high school: An ethnographic analysis. *American journal of Education*, 233-255.

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. *Psychological Science*. Publish

online April 23, 2014. Retrieved from http://pss.sagepub.com/content/early/2014/04/22/ 0956797614524581.

McMaster, K. L., & Campbell, H. (2008). New and existing curriculum-based writing measures: Technical features within and across grades. *School Psychology Review*, *37*(4), 550-566.

McMaster, K. L., Du, X., & Pétursdóttir, A. L. (2009). Technical features of curriculum-based measures for beginning writers. *Journal of Learning Disabilities*, *42*(1), 41-60.

McMaster, K. L., Du, X., Yeo, S., Deno, S. L., Parker, D., & Ellis, T. (2011). Curriculum-based measures of beginning writing: Technical features of the slope. *Exceptional Children*, *77*(2), 185-206.

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing a literature review. *The Journal of Special Education*, *41*(2), 68-84.

McMaster, K. L., Lembke, E., Brandes, D., Garman, C., Moore, K., Jung, P., & Janda, B. (2014). Data-based instruction in beginning writing: A manual. Unpublished manual, Department of Educational Psychology, University of Minnesota,cMinneapolis, U.S.

Morgan, P. L., Farkas, G., Hillemeier, M. M., Mattison, R., Maczuga, S., Li, H., & Cook, M. (2015). Minorities are disproportionately underrepresented in special education longitudinal evidence across five disability conditions. *Educational Researcher*, *44*(5), 278-292.

Murphy, A. F. (2009). Tracking the progress of English language learners. *Phi Delta Kappan*, *91*(3), 25.

National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the Education Success of Children and Youth Learning English: Promising Futures.* Washington, DC: The National Academies Press: DOI: 10.17226/24677.

National Center for Education Statistics. (2012). Data explorer. Retrieved from http://nces.ed.gov/nationsreportcard/naepdata/dataset.aspx

National Commission on Writing. (2003, April). *Writing: A ticket to work….or a ticket out: A survey of business leaders.* Retrieved March, 13, 2015 from http://www.collegeboard.com/prod_downloads/writingcom/writing-ticket-to-work.pdf

National Writing Project & Nagin, C. (2006). *Because Writing Matters: Improving Student Writing in Our Schools.* (2nd Ed), San Francisco, CA: Jossey-Bass.

Ovando, C. J. (2003). Bilingual education in the United States: Historical development and current issues. *Bilingual research journal*, *27*(1), 1-24.

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading &Writing Quarterly*, *19*(2), 139-158.

Pajares, F., Johnson, M. J., & Usher, E. L. (2007). Sources of writing self-efficacy beliefs of elementary, middle, and high school students. *Research in the Teaching of English*, 104-120.

Parker, C. E., Louie, J., & O'Dwyer, L. (2009). New Measures of English Language Proficiency and Their Relationship to Performance on Large-Scale Content

Assessments. Issues & Answers. REL 2009-No. 066. *Regional Educational Laboratory Northeast & Islands*.

Passel, J. S., Cohn, D., & Lopez, H. (2011). Hispanics account for more than half of nation's growth in past decade. Washington, DC: Pew Hispanic Center.

Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing writing*, *15*(1), 18-39.

Ritchey, K. D. (2006). Learning to write: Progress-monitoring tools for beginning and at-risk writers. *Teaching Exceptional Children*, 39,22-26.

Ritchey, K. D., & Coker, D. L. (2014). Identifying writing difficulties in first grade: An investigation of writing and reading measures. *Learning Disabilities Research & Practice*, *29*(2), 54-65.

Rolstad, K., Mahoney, K., & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational policy*, *19*(4), 572-594.

Romig, J.E., Therrien, W.J., & Lloyd, J.W. (2017). Meta-Analysis of criterion validity for curriculum-based measurement in written language. *The Journal of Special Education, 51 (2),* 72-82.

Rueda, R., Artiles, A. J., Salazar, J., & Higareda, I. (2002). An analysis of special education as a response to the diminished academic achievement of Chicano/Latino students: An update. *Chicano school failure and success: Past, present, and future*, 2, 310-332.

Samson, J.F. & Lesaux, N. (2015). Disadvantaged language minority students and their teachers: A national picture. *Teachers College Record,* 117, 1-26. Retrievd from: http://www.tcrecord.org.

Sandberg, K. L., & Reschly, A. L. (2010). English learners: Challenges in assessment and the promise of curriculum-based measurement. *Remedial and Special Education*. 32(2), 144-154.

Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, *18*(3), 337-354.

Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading & Writing Quarterly*, *23*(1), 7-25.

Silliman, E.R. (2014). Integrating oral and written language into a new practice model: Perspectives of an oral language researcher. In B. Arfe, J. Dockrell, & V. Berninger (Eds.), *Writing development in children with hearing loss, dyslexia, or oral language problems: Implications for assessment and instruction. (pp. 301-312).* New York, NY: Oxford University Press.

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *Tesol Quarterly*, 27(4), 657-677.

Slama, R. B. (2012). A longitudinal analysis of academic English proficiency outcomes for adolescent English language learners in the United States. *Journal of Educational Psychology*, *104*(2), 265-285.

Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research*, *75*(2), 247-284.

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. *The Cambridge handbook of literacy*, 112-133.

Soto, R. Ariel, G., Hooker, S., & Batalova, J. (2015). *States and Districts with the Highest Number and Share of English Language Learners.* Washington, D.C..: Migration Policy Institute. https://www.migrationpolicy.org/research/states-and-districts-highest-number-and-share-english-language-learners

Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research & Practice*, *15*(3), 128-134.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, *42*(8), 795-819.

Sullivan, A. L. (2011). Disproportionality in special education identification and placement of English language learners. *Exceptional Children*, 77(3), 317-334.

Sullivan, A. L., & Bal, A. (2013). Disproportionality in special education: Effects of individual and school variables on disability risk. *Exceptional Children*, 79(4), 475-494.

Stormont, M., Reinke, W. M., & Herman, K. C. (2010). Introduction to the special issue: Using prevention science to address mental health issues in schools. *Psychology in the Schools*, *47*(1), 1-4.

Sugai, G., & Horner, R.H. (2006). A promising approach for expanding and sustaining school wide positive behavior supports. *School Psychology Review, 35,* 246-259.

Wayman, M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, *41*(2), 85-120.

WIDA Consortium. (2013). *RtI 2: Developing a Culturally and Linguistically Responsive Approach to Response to Instruction & Intervention (RtI2) for English Language Learners*. Board of Regents of the University of Wisconsin System.

Wiese, A. M., & Garcia, E. E. (1998). The Bilingual Education Act: Language minority students and equal educational opportunity. *Bilingual Research Journal*, *22*(1), 1-18.

Wolf, M. K., Farnsworth, T., & Herman, J. (2008). Validity issues in assessing English language learners' language proficiency. *Educational Assessment*, *13*(2-3), 80-107.

Yanosky, T., Yen, S.J., Louguit, M., MacGregor, D., Zhang, Y., & Kenyon, D. (2011). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test, Series 200, 2008-2009 Administration* (WIDA Consortium Annual Technical Report No. 6.).

Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal*, *31*(4), 845-862.

Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome self-regulatory goals. *Journal of educational Psychology*, *91*(2), 241-250.

Zimmerman, B. J., & Kitsantas, A. (2007). A writer's discipline: The development of self-regulatory skill. In G. Rijlaarsdam (Series Ed.) and P. Boscolo & S. Hidi (Volume Eds.), Studies in Writing, Volume 19, *Writing and Motivation*, (51-69). Oxford: Elsevier.

APPENDIX

Teacher Consent Form

CONSENT FORM
Monitoring Motivation and Academic Growth in Writing for Young English Learners

Dear Teacher,

You are being invited to participate in a research study about young English Learner student's performance and growth in English writing. We are asking for your support with collecting student consent, helping with communication, scheduling with us to conduct short assessments with students, and completing short behavior/motivation scales for each student participant. Please read this form and contact us with any questions you may have.

This study is being conducted by R. Alex Smith, a PhD student, under the guidance of Dr. Erica Lembke, PhD, an Associate Professor of Special Education, from the Department of Special Education at the University of Missouri.

Background Information

The purpose of this study is to collect the information needed to help teachers identify young English Learners struggling in writing, create challenging academic goals, and monitor learning as compared to other English Learners with similar levels of English proficiency. The study will also collect information about how a student's motivation to write may change across time and influence academic achievement.

Procedures:

1. We will begin by contacting you to gather the following information: your schedule, preferred language of written communication home to your student's parent(s) or legal guardian(s), and schedule a time for a researcher to come disseminate letters of consent to students.
2. Students will have about two weeks to return their signed letters of consent and we ask that you periodically remind them to return the forms and communicate with us when letters have been returned.
3. Once consent has been returned, we will need to schedule times for assessment.
4. We will administer two forms of a spelling assessment and two forms of a sentence writing assessment in September, January, and May. We will also administer two forms of a story writing assessment in May. Each assessment takes 4-5 minutes. The sentence and story writing assessments can be administered to a whole group of students but the spelling assessment must be administered individually to each student. All assessments will be administered and scored by a researcher. Each student will spend about 20-30 minutes

completing these assessments once in the fall, once in the winter, and again in the spring.

5. You will be asked to complete a questionnaire about each participant's attitudes, behavior, and motivation toward writing at each time point. Each questionnaire will take about 1-3 minutes per student.
6. We will collect demographic data about your student, including age, sex, disability, race/ethnicity, district test scores, free/reduced lunch status, English proficiency, ACCESS test scores from the 2015/16 and 2016/2017 school years, migrant status, and native language.

Risks and Benefits of being in the Study

There are no known risks in participating in this study. However, your student will be practicing writing and the scored results will be provided to you for analysis. You will need to commit some time outside of your contract hours to complete the motivation questionnairres for each student. We will provide you with small gift cards for your time, not to exceed $30 total over the course of the study. The data from this study will help teachers in the future identify student strengths and areas of need in writing and create challenging writing goals based upon the performance of other young English Learners. The study will also provide important information about young English Learner's motivation to write in English.

Compensation:

Your students will receive small prizes, such as pencils or stickers, for completing writing assessments and returning letters of consent. You will receive a small gift card for your time.

Confidentiality:

The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your future relations with the University of Missouri or Columbia Public Schools. You may stop or withdraw from the study at any time without penalty or loss of benefits.

Contacts and Questions:

The researchers conducting this study Alex Smith and Erica Lembke. You may ask any questions you have now. If you have questions later, **you are encouraged** to contact Alex

Smith at 503-991-4135 or rashz9@mail.missouri.edu or Erica Lembke at 573-882-0434 or lembkee@missouri.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Campus Institutional Review Board, Office of Research, University of Missouri, 484 McReynolds Hall, University of Missouri, Columbia, MO 65211, *Phone:* 573-882-9585.

You will be given a copy of this information to keep for your records.

Statement of Consent:

I have read the above information. I have asked questions and have received answers.

__ Yes, I do consent to participate in the study.

__ No, I do NOT consent to participate in the study.

Your Name (please print): _____

Your Signature:_____ Date: _____

Signature of Investigator:_____ Date: _____

Parent Consent

CONSENT FORM

Monitoring Motivation and Academic Growth in Writing for Young English Learners

Dear Parent or Guardian,

Your child is invited to take part in a research study about young English Learner's writing in schools. We would like to use your child's writing scores and other information as part of our study. Please read this letter and contact us with any questions.

This study is being conducted by R. Alex Smith, a PhD student, and Dr. Erica Lembke, PhD, an Associate Professor of Special Education, from the Department of Special Education at the University of Missouri.

The purpose of this study is to collect information to help teach writing to young English Learner's and identify students that are struggling.

If you consent/would like for your child to take part in this study, he/she will be asked do the following:

1. Take two spelling tests and two sentence-writing tests in September, January, and May. Each test takes 4-5 minutes.
2. Write two short stories in May. Your child will write for 3 minutes for each story.
3. Your child's ESL teacher will answer some questions about how your child likes or does not like writing in September, January, and May.
4. We will also collect this information from the school: your child's age, sex/gender, if he/she has a disability, race/ethnicity, district test scores, ACCESS test scores from the 2015/16 and 2016/2017 school years, migrant status, and native language.
5.

Risks and Benefits of being in the Study

There are no known risks or direct benefits for being in this study. The information from this study will help teachers identify student strengths and areas of need in writing for other young English Learners in the future.

Confidentiality:

Your child's identifying information will be kept private. We will not include information that allows others to identify your child in any published paper. Research records will be stored securely and only approved researchers will have access to the records.

Voluntary Nature of the Study:

Participation in this study is voluntary. You do not have to include your child in this study if you do not want to and your or your child's relationship with the University of Missouri or Columbia Public Schools will not change. Your child may stop being in the study at any time for any reason without penalty or loss of any benefit.

Contacts and Questions:

If you have questions, **you are encouraged** to contact Alex Smith at 503-991-4135 or rashz9@mail.missouri.edu or Erica Lembke at 573-882-0434 or lembkee@missouri.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Campus Institutional Review Board, Office of Research, University of Missouri, 484 McReynolds Hall, University of Missouri, Columbia, MO 65211, *Phone:* 573-882-9585.

You will be given a copy of this information to keep for your records.

Statement of Consent:

I have read the above information. I have asked questions and have received answers.

__ Yes, I do consent for my child to participate in the study.

__ No, I do NOT consent for my child to participate in the study.

Child's name (please print):

_____

Parent/Guardian's name (please print):

_____

Signature of Parent:_____Date: _____

Signature of Investigator:_____Date: _____

**Flesh Kincaid reading ease of 63, grade level 7.3. However, score is slightly inflated because of formatting (inclusion of email addresses) and some technical vocabulary (e.g., participants, research, confidentiality). Content approaches 6th grade reading level.*

***This letter will be translated into the native language for parents who have indicated to the district that they would like communications home in their native language as well as according to feedback/guidance from the student's ESL teacher.*

Student Assent

ASSENT FORM

Monitoring Motivation and Academic Growth in Writing for Young English Learners

(This form is to be read aloud to each child participant.)

Hi! My name is _____. I am from the University of Missouri. I am working with your teacher this year. We are trying to find good ways to help kids with their writing.

If it's OK with you, I will ask you to write some words, sentences and stories. I'll give you some pictures or words to help you think of what to write. The activities will look like this (show sample prompts). I will also ask your teacher some questions about how much you like writing.

I will be looking at you ACCESS tests scores, MAP scores, your writing on the word, sentence and story assessments. I will also need information like your gender, age, native language, if you have a disability, and how much you have to pay for lunch.

If you don't want to write with me or share your information, you don't have to. Or, if you get tired of writing, you can stop at any time. Also, if you don't want your teacher to share information about your writing with me, you can just tell me or your teacher. No one will get mad at you if you don't want to write with me, or if you don't want your teacher to show me what you wrote. OK?

Do you have any questions? You can ask questions now, and if you think of questions later, you can ask them then, too.

Can you stop writing at any time? (should say 'yes')

Do you have to share your information with me if you do not want to? (should say 'no')

Is it OK if we do some writing activities together now?

Name:_____DATE:_____

Word Dictation

WD Form 16

Word List

1. had	16.wise
2. skin	17.chore
3. kiss	18.loop
4. met	19.slug
5. swam	20.self
6. desk	21.fate
7. bite	22.spout
8. lid	23.small

Picture Word

mouse

sun

cap

SAEBRS-AB

Impulsiveness	0	1	2	3

Academic Behavior				
Interest in academic topics	0	1	2	3
Preparedness for instruction	0	1	2	3
Production of acceptable work	0	1	2	3
Difficulty working independently	0	1	2	3
Distractedness	0	1	2	3
Academic engagement	0	1	2	3

Word Dictation Administration Directions

Directions for Administration of Word Dictation Task

Materials Needed:

1. Timer
2. Pencils
3. Directions for administration
4. Teacher copy of Word Dictation task
5. Student copy of Word Dictation task

Directions:

Say to the student: *Today we are going to do a writing activity. You will write some words for me. I will read each word two times, and then you will write the word on your paper. It's okay if you don't know how to spell a word. Do your best and then we will move on to the next word. Let's start with a practice word. Write the word "cat" on your paper. "Cat."*

Monitor the student to see that he/she is writing the word on the top line of his/her paper under "example". Don't worry about spelling mistakes. When the student is finished or pauses for more than 5 seconds on the practice word, demonstrate how to write the word on the line.

Now, you will write some more words. When you are finished with one word, move down a line and get ready for the next word. If you make a mistake, just cross it out. Do you have any questions? Remember to do your best! (Set timer for 3 minutes.)

*Here is your first word…*_____. Start timer after administering the first word.

Beginning with the first word, say each word two times, pausing briefly in between. Go on to the next word when the student is finished, or when the student pauses on a word for more than 5 seconds, in which case you would say to the student: "Try the next word." Do not provide any prompts to the student after the initial word reading. Read words at a consistent pace, without rushing the student. Time the student for 3 minutes. If the student finishes writing before 3 minutes, record the time remaining on the student form. If student is in the middle of writing a word when the timer rings, make a mark behind the last letter written before the timer rang, and score accordingly.

When the timer rings, say *Stop. Thank you for working so hard!*

Shortened Directions for progress monitoring:

Say: *Now will we write some words.*

I will say each word two times and you will write it. When you are finished writing a word, move down a line and get ready for the next word. Remember to do your best! (Set the timer for 3 minutes.)

When the timer rings, say: Stop. Thank you for working

Picture Word Administration Directions

Materials Needed:

1. Timer
2. Pencils
3. Directions for administration
4. Teacher copy of the task
5. Picture-word task for students

Directions:

Provide each student with a pencil and a picture-word prompt. Place the worksheet face up on the table in front of each student. Students should leave their pencils on their desks.

Say to the students:
Today we will do a writing activity. I will ask you to write some sentences. You will write one sentence for each picture in your packet. Keep your pencils down. First, let's name the picture on the front of your packet.

This is a car. (Point to the picture on the packet.)
What is this word? "car." (Make sure all students say the word.)
Let's make a sentence with this word. (Ask one or more students to make a sentence with this word.
What does a good sentence start with? (Prompt for capitalization.)
What does a good sentence end with? (Prompt for ending punctuation mark.) Choose one sentence to write on the board. Read this sentence aloud to the whole class.

You will write one sentence for each picture. (Point to the first item in the sample packet.)
Start at the top, then go down the page. Try to write a sentence for each picture. When you reach the end of a page, continue on to the next page. (Show the students with the sample copy).
If you reach the end before the time is up, go back and re-read your sentences and add details or more sentences.

Keep writing until the time is up and I ask you to stop. When I say "Stop", raise your hand with your pencil in it, like this. (Demonstrate.)

Remember to do your best work. If you don't know how to spell a word, just make your best guess. If you make a mistake, just cross it out.

Before we begin, let's read each word. Pencils should not be in your hand yet. Point to each word as I read it. (Read each word aloud to the students.)

Now, everyone turn back to the first page of your packet. You will have 3 minutes to write. Remember, this is not about finishing fast, this is about writing your best sentences. Do you have any questions?

Turn the page, pick up your pencils, and point your pencils to the first line. When I say "begin," write one sentence for each picture. Remember to do your best writing. Begin. (Start the timer).

Monitor participation. If individual students pause for about 10 seconds or say they are done before the 3 minutes have passed, say to the whole class: ***Keep writing until the timer rings.*** This prompt can be repeated if students should pause again. If students reach the last page before the end of the 3 minutes, say ***Go back and check your work or add more details***.

When the timer rings, say: ***Stop. Raise your hand with your pencil in it.***
Caution: When this is given to an entire classroom, sometimes students try to make it a competition to see who can finish first. If this happens, remind students at the completion of the task that it is NOT important to finish all of the sentences and that students who write really good sentences might take longer than students who write short sentences. We expect students to write really good sentences. Also, some students might be upset if they can't finish a sentence when the timer rings. Again try to reassure them that it's OK if they didn't finish.

Shortened Directions for Progress Monitoring:
Say: ***Do you remember how we did this before?*** (Point to an item in the sample packet)
You are going to write a good sentence for each picture. When you reach the end of a page, continue on to the next page. (Show the students what you mean with the sample copy).
Keep writing until I ask you to stop. Remember to do your best work. If you don't know how to spell a word, just make your best guess. If you make a mistake, just cross it out. Before we begin, let's read each word. Point to each word as I read it (Read each word aloud to the students. Make sure they follow along).
Now, turn back to the first page and point your pencil to the first line. When I say "begin", write one sentence for each picture. Make sure all the students are ready to start and say: ***Please begin writing*** (Start the timer set for 3 minutes).
Monitor students' participation. If individual students pause for about 10 seconds or say they are done before the 3 minutes have passed, say to the students: ***Keep writing until the timer rings.*** This prompt can be repeated if students should pause again. If students reach the stop page before the end of the 3 minutes, quickly mark the time on the stop page.

When the timer rings, say: ***Stop. Raise your hand with your pencil in it."***

Word Dictation Scoring Guide

Scoring Guide

CBM for Beginning Writers

Word Dictation (WD)

Materials:

1. Red and blue colored pencils: Blue = correct & Red = incorrect
2. List of administered words and student packet.
3. Record student name, week, and the date student completed the task.

Scoring Procedures:

For word dictation, count:

1. The number of words written (WW),
2. Words spelled correctly (WSC),
3. Correct letter sequences (CLS) , and
4. Incorrect letter sequences (ILS).

Words Written (WW)

1. Count the number of words written. A word is defined as a series of letters on a line or separated by spaces on each side.
    a. If the student is in the middle of writing a word when the timer stops, and they have written 2 or more letters, it counts as a word written.
    b. Score only the word that represents your best judgment of what the student meant to write for the target word

Words Spelled Correctly (WSC)

1. A word counts as a WSC only if it matches the target word. If the student spelled another English word but it does not match the target word, it is scored as an Incorrect Word (with the exception of homophones).
   Tip: Score the Word Dictation probes with the list of administered words next to the student response sheet to check answers.
   Example: target word is "drove" but student wrote "drive" (WSC = 0)
2. Underline incorrectly spelled words in red.
3. Calculate WSC by subtracting underlined words from WW.
4. Reversals of correct letter formation would cause the word to be scored as incorrect. Example: catɐ. (WSC = 0)

Correct Letter Sequences (CLS) and Incorrect Letter Sequences (ILS)

1. A correct letter sequence is one that contains any two adjacent, correctly placed letters.
2. Use the caret method for scoring. Place a blue caret ^ above two letters if it represents a correct letter sequence, and a red caret v below the letters if it represents an incorrect sequence. Score incorrect sequences first using a red pencil below the line. Then score correct sequences with a blue pencil above the line.
3. Score a correct letter sequence at the beginning of the word if the first letter of the word is correct. Score an incorrect letter sequence at the beginning of the word if the first letter is incorrect. Continue to score correct and incorrect sequences through the rest of the word. Score a correct sequence at the end of the word if the last letter is correct. Score an incorrect sequence at the end of the word if the last letter is incorrect.
4. If a word ends in a double letter (e.g., grass), and the student writes the word with only one letter, the sequence at the end of the word is scored with one incorrect letter sequence. The word would not count as a word spelled correctly. Consider the following examples (dictated word = grass):
   Example: _vr^a^sv

   (WW = 1, WSC = 0, CLS = 2, ILS = 2)

   Example: ^g^r^a^sv

   (WW = 1, WSC = 0, CLS = 4, ILS 1)

5. If a student omits a letter in the middle of a word, score with one incorrect letter sequence. Consider the following example where the student wrote wed for weed.
   Example: ^w^evd^

   (WW = 1, WSC = 0, CLS = 3, ILS = 1)

6. If a student doubles a letter inside a word, but otherwise has spelled the word correctly (e.g., classp for clasp), score an incorrect letter sequence on either side of the second double letter.
   Example: ^c^l^a^svsvp^

   (WW = 1, WSC = 0, CLS = 5, ILS = 2)

7. If student is in the middle of writing a word when the timer rings, score the letter sequences written up to the last letter. Do not score the final sequence as correct or incorrect.
   Example: ^c^l^a

   (WW = 1; WSC = 0; CLS = 3; ILS = 0)
8. Count correct sequences. Count incorrect sequences.

Picture Word Scoring Guide

Scoring Guide

CBM for Beginning Writers

Picture Word (PW) and Story Prompt (SP)

Materials:

4. Red and blue colored pencils: Blue = correct & Red = incorrect
5. Scoring Protocol and student packet.
6. Record student name, week, and the date student completed the task.

Advice to teachers for consistency:

You may choose to score critical skills more rigorously than described in this guide if that skill is an instructional focus. Be sure to score consistently from the beginning to the end of the year for all students. Remember: ***If you want to measure change, don't change the measure!***

General Scoring Procedures:

1. Read the sentence (PW) or entire writing sample (SP) first. Do your best to decipher what the student is writing. Sounding out what your student wrote may help in deciphering a word. For SP, also mark the beginning and end of each sentence using parsing guidance (see CWS/IWS directions below).
2. Count the number of words written (see WW directions below).
3. Underline incorrectly spelled words with red pencil as a spell checker would and calculate words spelled correctly (see WSC directions below).
4. Score and count correct and incorrect word sequences using the caret method: blue for correct ^ and red for incorrect ˅ (see CWS and IWS directions below).
5. Find Total Scores for each scoring procedure: WW, WSC, CWS, IWS

Words Written (WW)

1. Count the total number of words written, including all words spelled correctly and incorrectly. Ignore spacing problems unless the sample is very difficult to read (i.e., if you can distinguish between words even though they are close together, count them as individual words.)
2. If the student rewrites the story prompt on their paper, it counts towards their WW, WSC and word sequence scores.
3. If a student writes "The End" at the end of their story prompt, count it towards WW, WSC and word sequences. It does not need a punctuation mark to be counted as correct.
4. Additional guidance:

a. For repeated words: Count the first 3 words and cross out the rest.
Example: It was fun fun fun ~~fun fun~~.

(WW=5)

b. Hyphenated words are counted as one written word.
Example: He is a well-known actor.

(WW=5)

c. Whether or not there is punctuation after the last correct word, include the last word in WW score.
Example: Sally went to the store

(WW = 5)

d. When the timer rings, if there is no punctuation after the last word AND the student has written two or more letters, it is counted as a word written.
Example: Sally went to the stor

(WW = 5)

Words Spelled Correctly (WSC)

1. Underline incorrectly spelled words in red. Score these words the same as a spell checker would.
2. Additional guidance:
   a. Words that are spelled correctly (even if they do not make sense) should NOT be underlined.
   Example: Sally went two the store.

   (WW = 5, WSC=5)

Note: The word ‘‘two’ is *used* incorrectly but is not underlined because it is *spelled* correctly.

   b. Words with reversals are incorrect (i.e., “b” for “d”)
   Example: The vduddlev floated up and popped.

   (WW=6, WSC=5)

   Example: I like vcatav.

   (WW=3, WSC=2)

3. The following are counted as words spelled correctly (NOT underlined):
   a. Acronyms that are capitalized
   Example: Both “TV” or “T.V.” can be scored as correct.

b. Abbreviations of proper nouns with correct capitalization and punctuation
Example: "P.E." for Physical Education
b. Other common abbreviations
Examples: min, lb, hr, etc.
d. Numbers used *correctly,* including *dates used correctly*
Example: I had 4 toys. My birthday is 8/04/99. For both examples, **WW=4.**

> *Note:* Numbers used in place of words are incorrect (although numbers are counted as words in WW). *Example*: We went 2 the pool.
>
> **(WW= 5, WSC=4)**

e. Symbols used in place of words, such as "&" used for "and"
f. Videogames or other popular culture titles (Minecraft, Pogo), even if they are not found in the dictionary

4. Calculate Words Spelled Correctly (WSC) by subtracting the number of underlined words from the number of Words Written (WW).

Correct Word Sequence (CWS) and Incorrect Word Sequence (IWS)

1. For PW, the target word does NOT need to be used in the sentence written.
2. Students are not penalized for starting their story with a sentence that does not make sense in the context of the given prompt ("It was the last day of school so I decided to…I would go outside."). Score the first sentence as you would any other sentence.
3. Parsing: Place a vertical blue line at the beginning and end of each sentence (you may have to judge where the sentence should end).
Example: |Sally went to the stoer |she bought some chocolate and mashmellows to make

desart. |

4. Using and: If a student uses *and* more than twice in a sentence and the sentence has more than two clauses, break up the sentence appropriately so that only two *and's* are in each sentence (see example below). *And* may not be used at the beginning of a sentence.
5. Use the caret method for scoring word sequences. Place a red caret below both sides of an incorrectly *spelled* or *used* word, indicating incorrect word sequences (IWS). If two words in a row are spelled and used correctly, place a blue caret *above and between* the two words indicating a correct word sequence (CWS).

> ***Scored example:***
>
> | ^We ^went^ to^ the^ store ^and ^went ^home^ and^ then ^went^ to^ the^ park$_{v}$|
>
> $_{v}$and$_{v}$ met^ my ^friends^and ^they^ were^ excited$_{v}$| $_{v}$and^ we^ played ^soccer^and
>
> ^we ^had ^fun^.|

6. At the beginning of a sentence, the initial word sequence is correct if both of the following are true:

   a. the first word is correct.
   b. the first letter is a capital.

7. Capital letters:

   a. The first letter of the sentence must be capitalized, or the sequence where the word should be capitalized is incorrect. For example, if the first word is "The" and it is spelled correctly but not capitalized (the), it would be scored like this: vthe^ boy...
   b. Other capital letters within the sentence should be ignored.
   c. Proper nouns need to be capitalized.
   d. "I" must be capitalized to be counted as a word spelled correctly. A lowercase "I" results in two incorrect word sequences whether within or at the beginning of sentences.
      Example: v i v like ^ to ^ swim ^.

      (WW = 4, WSC = 3, CWS = 3, IWS = 2)

   e. For Story Prompts, if the student begins the writing sample as a continuation of the story prompt, do not penalize if they do not capitalize the first letter of their writing sample.
      Example: [It was the last day of school so I decided to…] |^go^ outside^ and^ play^.| **(WW = 4, WSC = 4, CWS = 5, IWS = 0)**

> ***Note:*** If upper and lower case letters look similar and it is difficult to tell whether the letter is an upper or lower case letter, ignore capitalization. These letters include: p, s, o, t, c, u v, w, x, y, z.

8. At the end of a sentence the final word sequence is correct if *both* of the following are true:

   a. the word is correct

   b. the sentence ends with correct punctuation

Final words that are spelled correctly and make sense in context are correct words.

Example: ^I ^love ^summer^. | v Its v my ^favorite^ season^.

Final words that end without punctuation or with incorrect punctuation create an IWS, so

place an inverted red caret between the last word and the incorrect (or missing) punctuation.

Example: ^I^ love^ summer^ because^ it^ is ^warm ^and ^sunny ˅? | ^It^ is ^my^ favorite ^season^.

<u>If the last sentence does not include punctuation</u>, and the student stopped writing because of the time limit, leave the last sequence unscored.

9. If a student attempts to write a quote, as clearly marked by a "he/she/I said, "___", correct

   capitalization and punctuation are necessary in order to mark CWS.

   Example with incorrect capitalization and punctuation:

   ^She ^ said, ˅ did ^ you^ have ^a^ great^ day? ˅ |^ I^ said ˅ no. ˅

   Explanation:

   a. An IWS caret is between "said" and "did," because there is no quotation mark and "did" is not capitalized.
   b. An IWS caret is between "day" and the beginning of the next sentence because there is no closing quotation mark.
   c. IWS carets are on either side of "no" because there is not a comma after said, a quotation mark, nor a capital N.

   Example with correct capitalization and punctuation:
   ^She ^ said, ^ "Did ^you ^have ^a ^great ^day?" ^ | ^ I ^ said, ^ "No." ^

   Commas must be used correctly in dialogue such as: Dad said, "Go to your room."

   Ignore all other commas (either when missing or used incorrectly).

10. To assign CWS or IWS for verb tense shifts, follow these guidelines:

    a. If a student switches tense multiple times, either stick with the first tense used by the student, or apply the "majority of tenses" rule (i.e., count the number of verbs and verb tenses in the passage; verb tenses in the majority are scored as CWS and those not are scored as IWS).
    b. In cases where student uses an equal number of more than one tense, stick with the first tense used.
    c. Indirect quotations (no quotation marks necessary) permit verb tense shifts, as long as grammatically correct:

Example: ^She ^told ^me^ a ^lot ^of ^my ^friends ^are ^coming ^over ^because ^it's ^my ^birthday! ^

11. If there is a missing word, score as one incorrect word sequence.

Example: ^The ^ fish v blue ^ and ^ green ^.

Example: ^My^ window v.

12. Compound words that are written as two words are incorrect. e.g. "home work" should be written as "homework." This would be three IWS's, but the words are counted as correctly spelled words.

Example: ^ "All v home v work v must^ be^ turned ^in ^by ^Wednesday," ^the ^teacher ^said^.

13. Apostrophes must be used correctly in contractions and to show possession. If an apostrophe or contraction is used incorrectly, score as incorrect

Example: ^I^eat v egg's v in^the^morning^.

14. Colloquialisms: writing should reflect the conventions of standard English. Consider the following example:

Example: v Awwwwwwwwwww v a ^ bug. ^

(WW = 3, WSC = 2, CWS = 2, IWS = 2)

Try scoring this sample by yourself:

The cave was very dark. I try to close my eyes, so I couldn't see anything, but that didn't help. Than I hear some one breathing. I try to stream but nother came out. The breathing became close and close to me, and the worst Part was that I couldn't see athing. At first I thought meself that I an Just emaging stuff.

Fidelity of Administration for Word Dictation

Accuracy of Implementation Rating Scale (AIRS)--

CBM-W: Word Dictation

Implementer:	Observer/rater:
Date:	
Start time:	End time:

Part I. Administering the Assessment. Observe the assessment implementation, complete the checklist to the extent that the components were administered, and write detailed notes regarding other components observed.

	Yes 1	No 0	N/A	Observation notes:
1. Has materials on hand				
a. Timer				
b. Pencils				
c. Directions for administration				
d. Teacher copy of the task				
e. Word dictation task for students				
2. Follows the directions in order				
a. Places student copy in front of student				
b. Explains what to do if student does not know how to spell a word				
c. Reminds student to do his/her best work				
d. Practices a sample word				
e. Demonstrates how the student should proceed through task				
f. When a student pauses on a word for more than 5 seconds,				

says to the student: "Let's go on to the next word."				
3. Overall demonstration skills				
a. Reads directions accurately				
b. Demonstrates by pointing when appropriate				
c. Pause for questions				
4. Timing				
a. Says "Here is your first word."				
b. Starts/stops timer at the correct times				
c. Times student for 3 minutes				
d. Says "Stop. Thank you for working so hard."				
e. Marks administrator copy as needed				

Fidelity of Administration for Picture Word

Accuracy of Implementation Rating Scale (AIRS)--

CBM-W: Picture Word Prompt

Implementer:	Observer/rater:
Date:	
Start time:	End time:

Part I. Administering the Assessment. Observe the assessment implementation, complete the checklist to the extent that the components were administered, and write detailed notes regarding other components observed.

	Yes 1	No 0	N/A	Observation notes:
5. Has materials on hand				
a. Timer				
b. Pencils				
c. Directions for administration				
d. Teacher copy of the task				
e. Picture-word task for students				
6. Follows the directions in order				
a. Places prompt in front of each student				
b. Presents an example of a Picture-word prompt				
c. Demonstrates how students should complete the entire Picture-word task with the sample copy				
d. Reminds students to do their best work				
e. Demonstrates how to deal with spelling difficulties while taking test				

f. Reads each word on the Picture-word task				
g. Prompts students to continue working until the timer rings, if necessary				
7. Overall demonstration skills:				
a. Reads directions accurately				
b. Demonstrates by pointing when appropriate				
c. Pauses for questions				
8. Timing				
a. Says "Please begin writing"				
b. Starts/stops timer at the correct times				
c. Times students for 3 minutes				
d. Says "Stop. Raise your hand with your pencil in it."				
e. Marks administrator copy as needed				

VITA

R. Alex Smith was born in Oxford, Mississippi and predominately raised in Batesville, Mississippi. He received his B.A. in Psychology with a minor in English from the University of Mississippi in 2002. He received his M.Ed. in Special Education with an emphasis in Severe and Profound Intellectual Disabilities from the University of Georgia in 2007. He completed his Ph.D. in Special Education with an emphasis in Learning Disabilities from the University of Missouri-Columbia in 2018.