Spring 2020

Curriculum Based Measures for Screening English Language Learners: What We Know and Future Directions

Colleen H. Robinson
Western Kentucky University, colleenr1994@gmail.com

CURRICULUM BASED MEASURES FOR SCREENING ENGLISH LANGUAGE LEARNERS: WHAT WE KNOW AND FUTURE DIRECTIONS

A Specialist Project
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Specialist in Education

By
Colleen Robinson

May 2020

CURRICULUM BASED MEASURES FOR SCREENING ENGLISH LANGUAGE LEARNERS: WHAT WE KNOW AND FUTURE DIRECTIONS

Date Recommended 4/15/2020

Sarah Ochs, PhD, Director of Specialist Project

Carl Myers, PhD

Joseph Cangemi, PhD

Cheryl D Davis Digitally signed by Cheryl D Davis Date: 2020.05.04 12:29:55 -05'00'

Dean, The Graduate School Date

ACKNOWLEDGMENTS

First, I would like to thank the Lord for providing me the opportunity to help students, parents and teachers to do and be their best. I would like to thank my director of my specialist project, mentor, and friend, Dr. Ochs for guiding me through this process. Graduate school would not be the same without her guidance. I would also like to thank Dr. Cangemi, for being a source of wisdom throughout my two years at Western Kentucky University. I will always cherish our deep and meaningful conversations, laughs, and coffee times we shared. Finally, I would like to thank my family and friends for the continuous support through my time in graduate school. I would not be able to do it without you all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CURRICULUM BASED MEASURES FOR SCREENING ENGLISH LANGUAGE LEARNERS: WHAT WE KNOW AND FUTURE DIRECTIONS

Colleen Robinson May 2020 33 pages

Directed by: Dr. Sarah Ochs, Dr. Carl Myers, Dr. Joseph Cangemi

Department of Psychology Western Kentucky University

Curriculum based measures (CBMs) are used to assess students' academic achievement by screening for risk and monitoring progress toward a goal. English Language Learners (ELLs) may acquire academic skills in a different way and it is important to investigate whether CBMs function in the same way for this diverse subset of our student population. The purpose of this paper was to review current literature about using CBM to screen ELL students. Ten articles were found which discussed the use of CBMs with ELL students. More specifically, these articles discuss the use of reading and written expression CBMs as valid measures in predicting ELL student's performance on high-stakes achievement tests. Limitations and future directions are discussed.

Introduction

Development of CBM

Curriculum-based measures (CBMs) are quick and easy to administer assessment tools used to assess academic performance. About forty years ago, Dr. Stanley Deno, the creator of CBMs, researched the best way for special education teachers to track student progress or response to curriculum. This was to aid in making changes to instruction and increasing the likelihood of success (Deno, 1985). After evaluating the validity and reliability of what was currently being used, Deno found the current ways of evaluating student performance, specifically for students with disabilities, were not accurately measuring students' abilities (Deno, 1985). Deno and his research team found relying on teachers as reporters or referrers for special education was not a reliable or valid method. Deno wanted to create a better measurement system that was reliable, valid, and simple to administer and interpret. Thus, Deno and his team developed CBMs as a direct measure of academic behavior in the areas of reading, spelling, and writing. These measures are standardized and are an observation of academic skill performance. CBMs were found to yield reliable and valid data that were a useful indicator of academic achievement and related to state test performance (Deno, 1985). These data were then used to support educational decision making in schools for various reasons including screening for risk, assisting with special education referrals, developing individualized educational plans (IEPs), monitoring student progress, and planning for modifications to instruction or intervention (Deno, 1985).

CBMs use frequent and direct observation of academic skills to measure students "academic health" and how well the student is responding to the instruction being provided (Deno, 1985; McMaster et al., 2017) CBM probes were developed to match each grade level, approximate ability, and molded around the academic curriculum to accurately represent what students are being taught in school (Hosp, Hosp, & Howell, 2007). This allows teachers and other educational professionals to get an idea of how well students are understanding material and better plan modifications to instruction or develop interventions. Extending beyond Deno's initial development, CBMs are now available to assess reading fluency and comprehension, mathematic computation and concepts and applications, early literacy, early numeracy, spelling, and written expression.

Screening and Progress Monitoring

Since CBMs are used to assess students' academic abilities, they are a useful tool in creating educational goals and modifications to instruction or intervention plans (Hosp et al., 2007). CBMs can be used to screen student risk and progress monitor academic skills. CBMs are a criterion referenced measure. This means they are used to see if students have reached certain levels of academic performance, making it easier to determine if a student has or has not mastered a skill (Hosp et al., 2007). This is also known as mastery measurement. CBMs help to accurately identify specific skill deficits, prepare appropriate, matched academic interventions, and to progress monitor students' academic skills (Hosp et al., 2007).

As such, CBMs fit well within the Response To Intervention (RTI) model (Hart & Stebick, 2016). The National Center on Response to Intervention (NCRTI, 2010) outlines

RTI to have four main components: (a) schoolwide, multilevel instructional and behavioral supports; (b) universal screening; (c) progress monitoring; and (d) data-based decision making for instruction and disability identification. RTI is a three-tiered model which incorporates the use of varying levels of intensive evidence-based intervention and instruction to meet students learning needs. Through the three-tiered RTI model, data are collected through screening and progress monitoring, which is frequently conducted using CBMs (Nese, Park, Alonzo, & Tindal, 2011). Tier 1 of RTI is the general education setting or the "everyday" curriculum. If a student does not adequately respond to the level of services provided to all students in Tier 1, they may require more support and "move up" to Tier 2. Tier 2 provides students with supplemental instruction in a certain academic area because they did not respond to the Tier 1 instruction. These services may include additional instructional time, small group work, or interventions. Tier 3 consists of more individualized and intensive support for students who did not make satisfactory progress even with Tier 2 services. If students are making insufficient progress within their Tier 3 intervention, this will typically lead to a special education referral. The screening and progress monitoring used to measure a student's progress is the piece which directly affects student movement across tiers. Thus, the use of CBMs within the RTI process are extremely valuable.

Within RTI, CBMs are used for two reasons: Screening and progress monitoring. For screening, all students within a grade complete a CBM in an academic area (typically reading and sometimes math). This score is used as a quick temperature check or snapshot of performance to quickly identify any students who may be at risk. For progress monitoring, CBM can be used to track the growth of students within an

academic area or measure the impact of an intervention on performance. CBM can also be used class-wide to evaluate the instruction the teacher is providing. Because they are so quick and simple to administer, CBMs are particularly useful for measuring academic growth over time (VanDerHeyden & Burns, 2005). With the simplicity of CBM, these measures can be administered frequently, often weekly, to see if interventions are or are not helping the student improve.

VanDerHeyden and Burns (2005) researched the use of CBM data to monitor the progress of students with math instruction. Students completed CBMs daily and data were used to track progress in order to see how they responded to the math instruction. Results found an increase of improved performance from 38% children in the frustrational level in January to only 24% in April (VanDerHeyden & Burns, 2005). This showed using CBM to track progress is a useful way to collect academic data.

Frequently, educators and interventionists measure and interpret rate of improvement (ROI) when making educational decisions. ROI is used to measure a student's current or initial performance and based on their progress (or growth), the time it will take for the student to improve and reach expected, grade level performance. CBM data are frequently used to calculate ROI. By doing this, educators use data to make decisions about instruction and see if students need a more or less intensive intervention.

CBM Standard Procedures

CBMs have a standard set of administration and scoring procedures which are similar across skill areas (Deno, 2003). For most administrations, the examiner reads a standard set of instructions to the student about how to complete the task. For example, for reading fluency, the examiner explains to the student that they will be timed and

asked to read a passage out loud for one minute. Then, CBM scores are typically rate-based or calculated by how many items (words, letters, problems, digits) a student gets correct per minute (or other brief period of time). In the reading example, the examiner will follow along with a copy of the passage, while the student reads, to record errors and calculate how many words he or she read in that minute. In written expression, a student's writing sample is scored using measures such as total words written (TWW) or words spelled correctly (WSC). With this information, the scores can be compared to local, national, or previous norms to see where how the child's reading fluency compares to peers. Based on these scores, a student could be performing below, at, or above grade level. These results are useful for screening and progress monitoring decisions regarding instructional need or risk.

Validity of CBM

Beginning with Deno's early work, many studies have examined CBMs relationship to state test performance and use for identifying students who are at risk of failing statewide achievement tests (e.g., Buck & Torgesen, 2003; Grapin, Kranzler, Waldron, Joyce-Beaulieu, & Algina, 2017; Silberglitt & Hintze, 2005; Wood, 2006; Yeo, 2010). This research has looked at the validity of CBM to determine if these measures can predict performance on high stakes tests, even for students with disabilities and English Language Learners (ELL). Yeo (2010) conducted a meta-analysis on the predictive validity of reading CBM (R-CBM) and statewide achievement tests. He found a large correlation between R-CBM and results of statewide tests (Yeo, 2010). These results found a fairly large correlation of .69 ($p < .05$), suggesting R-CBM is a valid predictor of performance on statewide reading tests.

English Language Learners

English Language Learners (ELLs) are students who are limited in English language proficiency (Reed, 2013) or who are in the process of learning English (Unruh & McKellar, 2017). The National Center for Education Statistics (2018) found that between 1 and 20% of schools within the United States are made up of students who are ELLs. Because some states have such a large population of ELLs, it is important that schools ensure programs help students reach English proficiency. Spanish is the primary language of almost 4 million ELL students in the U.S. as of Fall, 2015.

According to the U.S. Census, ELL students between the ages of 5 and 17 have doubled between 2000 and 2010. In the Fall of 2015, there were 4.8 million ELL students nationally, making up 9.7 percent of the U.S. public school's population (McFarland et al., 2018). Mastering language is the single most important predictor for academic success for ELL students (Reed, 2013). Research has found that it takes about five to seven years to master academic English. This means even though students are able to talk in social contexts, they may struggle to perform academically.

Based on this information, it is no surprise ELL students are frequently presented with challenges in school that make them more susceptible to being placed in special education. In some instances, this may be a fitting placement, but other times this is an error due to the effect of language deficits.

This diverse group of students is a population that is continuing to grow (Rhodes, Ochoa, & Oritz, 2005). This requires schools to adapt and provide measures that are reliable and valid for educational decision making for ELL students. Trying to find a valid, useful measure for ELL students can be extremely challenging (Unruh &

McKeller, 2017). Many academic tests used in education do not include ELLs in their norming sample, there are few adequate tests that can be used to determine whether an ELL is struggling academically because of a language deficit or because of an academic deficit (Unruh & McKeller, 2017). Additionally, ELL students are at an increased risk of academic failure, necessitating accurate universal screenings to identify risk and implement targeted intervention strategies.

When understanding how screeners work for ELLs, it is important to have a brief understanding of language acquisition. Students acquiring language are across a spectrum of proficiency. Students who are acquiring another language vary in the amount of time it can take to be able to speak, understand, and socially and academically comprehend a new language (Unruh & McKeller, 2017).

There are six stages of second-language acquisition (Robertson & Ford, 2009). The first stage is the pre-production period, also known as the silent period. During this time, which is usually about six weeks or so, students are observing and listening (Robertson & Ford, 2009). Stage two is early production. In this phase, students are beginning to use a few words and small sentences, but errors are frequent. Stage three is called speech emergent. In this stage, speech is starting to become more frequent, and errors are beginning to decrease, while new and improved language is increasing. Stage four is beginning fluency. In this stage, students are beginning to become more fluent and able to talk in social situations. Beginning fluency is also the stage that academics begins to come into play. Academic understanding is challenging for the student, but they are progressing. The next stage is intermediate fluency. In intermediate fluency, students have the ability to understand, fluently learn and think in the second language. The final

stage is advanced fluency. At this stage, the student is able to understand new information and think complexly in their second language. Errors still occasionally occur, but are rare (Robertson & Ford, 2009). An important note about second-language acquisition is that social understanding and language comes before academic skills and understanding. Students are able to talk with their peers on the playground and have conversations with their friends before they are able to interpret and understand academic material (Robertson & Ford, 2009). Proficiency in academic language can take between five to seven years (Collier, 1989; Cummins, 1981; Mohr & Mohr, 2007).

This can be frustrating for teachers. A study was conducted to look at teacher's opinions on having ELL students in their classroom (Walker, Shafer, & Iiams, 2004). Based on these surveys, 70% of the teachers were not interested in having ELL students in their classroom and 87% of the teachers said they did not receive professional development about teaching ELL students (Walker et al., 2004). These results highlight additional challenges that ELL students face in school.

Second-language acquisition is a fairly lengthy process that is challenging, especially for students who are trying to manage learning a new language and demonstrate proficiency in academic skills. Because this is a prolonged process, ELLs may acquire skills differently than students who are native English speakers.

Purpose of Study

There is a significant body of research supporting the use of CBM to make decisions about students in schools, however, there is less guidance on the use of CBM with ELLs. Given the increasing number of culturally and linguistically diverse students in our schools, it is important to understand the scope of the research using CBM with ELLs.

The purpose of this systematic review was to gain more knowledge about the utility of CBM as an academic screener for ELLs. Both progress monitoring and universal screening are important, but because universal screening is designed to identify students at risk, this review will focus on the literature on CBM and universal screeners for ELL students. Language is a primary skill and necessary for the further development of academic skills. This specialist project was a systematic review of the literature on the use of CBM for screening with ELLs. Specifically, a review for studies that examined the use of CBMs for screening with ELLs to predict future performance on state tests.

Method

Procedures

This project was a systematic review of the literature on curriculum-based measures as universal screeners with English Language Learners. Electronic search of the literature occurred within the following databases: Academic Search Complete, Education Full Text (EBSCOhost), ERIC (ProQuest), ERIC (EBSCOhost), PsycARTICLES (EBSCOhost), and PsycINFO (EBSCOhost). The search terms used were curriculum based measurement/"CBM", English language learner/"ELL", screener/universal screener. The Preferred Reporting Items for Systematic Reviews and Meta-analysis Model (PRISMA; Moher et al., 2009) was used to structure the inclusion of articles and illustrate each stage of the search. Figure 1 includes the number of articles included and excluded at each stage of the PRISMA model. PRISMA is an evidence-based method for searching and reviewing research (Moher et al., 2009). Only full-text, peer-reviewed articles and dissertations published between 1990 and 2019 were included. Articles also needed to examine the validity of CBMs for predicting state test performance. After initially identifying articles, abstracts were screened and excluded due to either not being an empirical study or not having relevant information regarding the predictive validity of CBMs on high-stakes tests. Full text of the remaining 23 articles was reviewed and an additional 13 articles were excluded because they did not examine predictive validity or were not an empirical study. Then, the final eight articles were comprehensively reviewed and summarized.

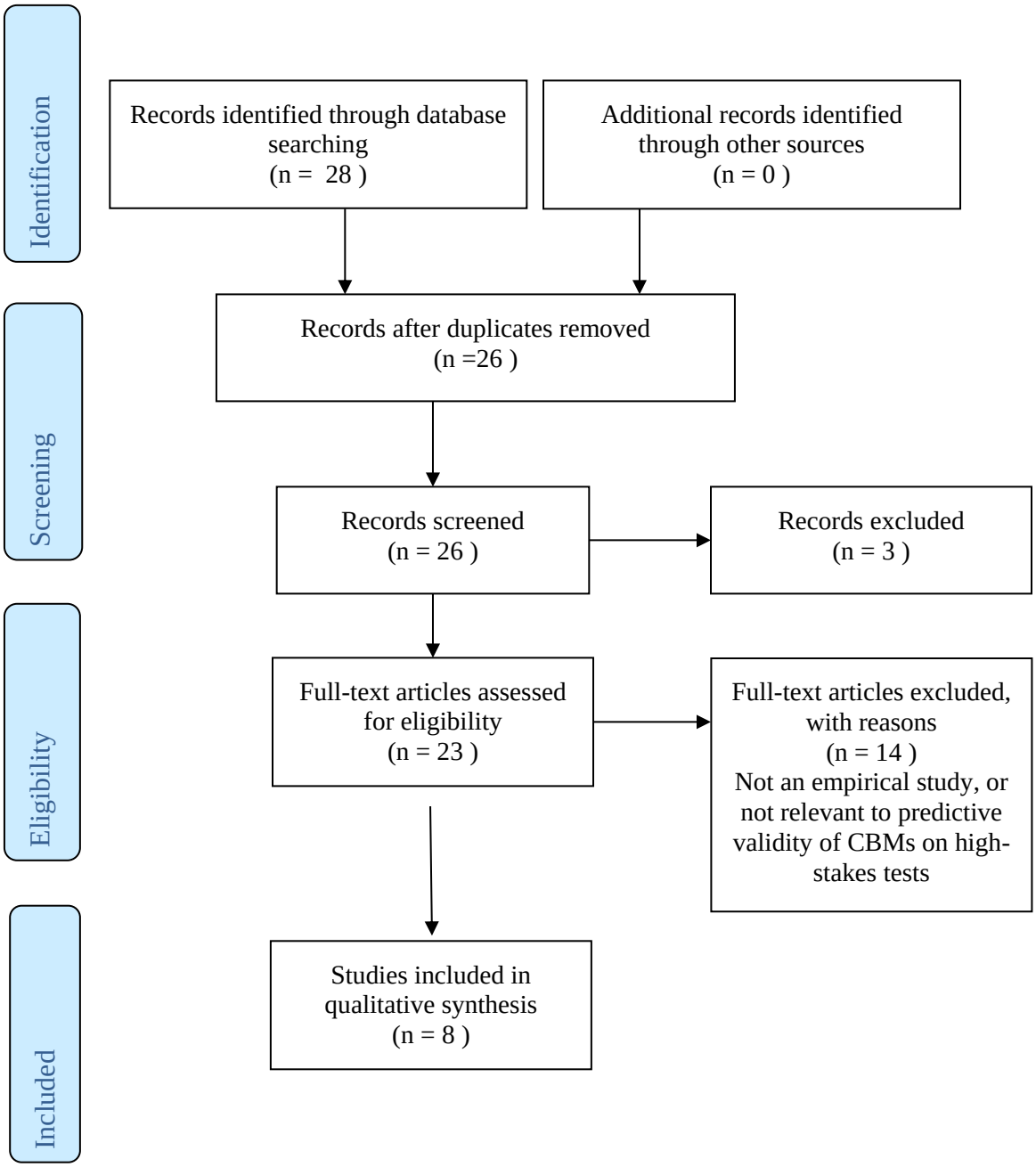

Figure 1: PRISMA 2009 Flow Diagram (Moher et al., 2009)

Results

Of the 28 articles initially identified in the search, eight were found to meet all inclusionary criteria and were included in this systematic review. Results were organized by academic content area and of the eight articles included, five were in the area of reading, and three were in written expression. Articles are listed in Table 1.

Reading

Domínguez de Ramirez and Shapiro (2006) examined Oral Reading Fluency (ORF) growth among Spanish-speaking ELLs in both English and their native language. The following research questions were assessed: Did ELL students in general education have higher reading scores in English than the ELL students in a bilingual education setting? Did students in the general education setting have higher rates of growth over the year in reading English than the ELL students in a bilingual setting reading Spanish? How did the growth over the year compare for the ELL students regarding speaking Spanish and English?

The sample included 695 students in grades K through 5 in bilingual and general education settings, with 165 of these students receiving bilingual education (48% females, 52% males). The 165 ELL students were primarily from a Mexican American community. Given the nature of the bilingual education program, it is hard to determine the degree of English proficiency of the students, however, it can be inferred the students in the sample were not fully competent in English (Domínguez de Ramirez & Shapiro, 2006). The ELL students attended a transitional bilingual education program which consisted of core subjects being taught in Spanish, while providing opportunities for learning English literacy skills tied in with the curriculum.

Table 1

Study Characteristics

Study	n	Grade(s)	CBM Type	Criterion
Domínguez de Ramirez and Shapiro (2006)	695	K-5	CBM-ORF	Texas Assessment of Academic Skills; Developmental Reading Assessment
Vanderwood, Linklater, and Healy (2008)	280	1	R-CBM NWF Maze	California Achievement Test, Sixth Edition
Muyskens, Bets, Lau, and Marston (2009)	1,205	5	*Invitation to Literacy* CBM-ORF	Minnesota Comprehensive Assessment
Richardson, Hawken, and Kircher (2012)	719	4-6	AIMSWeb Maze	English Language Arts Criterion Referenced Test
Kim, Vanderwood, and Lee (2016)	522	3	DIBELS ORF, DIBELS Maze	California Standards Test-English Language Arts
Wallace, Campbell, Lembke, Long, and Ticha (2008)	183	9-12	WE-CBM	Minnesota Basic Standards Test/Minnesota Comprehensive Assessment
Campbell, Espin, and McMaster (2013)	36	10-12	WE-CBM	Test of Written Language-3rd Edition; Test of Emerging Academic English; Minnesota Basic Skills Test
Keller-Marguilis, Payan, Jaspers, and Brewton (2016)	50	4	WE-CBM	State of Texas Assessments of Academic Readiness

Note. CBM = Curriculum-based measure, ORF = Oral Reading Fluency, R-CBM = Reading Curriculum-based measure, NWF = Nonsense word fluency, DIBELS = Dynamic Indicators of Basic Early Literacy Skills, WE = Written Expression

The Texas Assessment of Academic Skills (TAAS; Texas Education Agency, 1997) was used to measure grades 3-5 reading scores and the Developmental Reading Assessment (DRA) was used to measure grades 1 and 2. R-CBM was used to measure oral reading fluency in the fall, winter, and spring. Both English and Spanish passages were used and were selected based on each student's instructional level in reading. By gathering this information, two groups were identified: students who met district reading standards and students who did not. Researchers then made a comparison between the performances of the English-speaking general education students and the ELL students in the bilingual classroom.

Results indicated that general education English-speaking students performed better than the ELL students in the bilingual classroom. Results revealed all students (ELL and non-ELL) showed significant growth in reading. The group main effect was significant [$F(2, 30) = 109.88$, $p < .001$, $n^2 = 0.489$], which suggests Spanish-speaking ELLs read less fluently in Spanish than the general education students did in English.

The interaction between curriculum type and grade was not significant. An analysis of variance of bilingual students reading in English versus Spanish found a significant main effect for type of language [$F(1, 52) = 5.22$, Wilks lambda = 0.91, $p < .05$, $n^2 = 0.09$] and a significant interaction between type of language and grade [$F(4, 52) = 14.89$, Wilks lambda = 0.47, $p < .001$, $n^2 = 0.53$]. This study offers some valuable information regarding the performance of Spanish-speaking ELL students, especially those who are in the process of acquiring English (Domínguez de Ramirez & Shapiro, 2006). Results of this study demonstrate that CBMs can be sensitive to assessing

language and literacy development of Spanish ELL students (Domínguez de Ramirez & Shapiro, 2006).

Vanderwood, Linklater, and Healy (2008) examined Nonsense Word Fluency (NWF) used with ELL students. Specifically, the purpose of the research was to see how NWF with first grade ELL students was related to their performance on third-grade English literacy outcomes including R-CBM, Maze, and the state testing in reading. The study sample included 280 first-grade students from an urban elementary school in southern California with 90% ELL (Hispanic background) and 100% of the students receiving free or reduced-cost lunch. The students were originally assessed in first grade and then again in third grade. The students in the sample took the California English Language Development Test (CELDT) and around 80% were classified as beginning level, 13% were in the intermediate and 8% were advanced.

Students' NWF and Stanford Achievement Test, Ninth Edition (SAT9) was assessed at the beginning of first grade. The California Achievement Test, Sixth Edition (CAT6), R-CBM, and Maze was used to assess the students in the spring of third-grade. Results from NWF and the SAT9 indicated students from the beginning of first-grade to the spring show growth of over 36 sounds with a growth rate of 1.43 sounds per week. Overall, NWF was significant in predicting the students outcomes on their third-grade Maze, R-CBM and CAT-6 ($p < .01$). Using NWF to screen and identify ELL students who need additional services is considered to have initial empirical support (Vanderwood et al., 2008).

In Minneapolis Public Schools, Muyskens, Bets, Lau, and Marston (2009) investigated the predictive validity of oral reading fluency CBMs to determine if they

were an appropriate, valid measure for students who are ELLs. Specifically, Muyskens et al. (2009) examined scores on an oral reading fluency CBM as a predictor of later reading performance on state tests for ELL students (N = 1,205). Students came from three distinct ELL populations: Spanish, Hmong, and Somali. Students from all three group were frequently grouped together for instruction despite the vast language differences. The participants were fifth-grade students who had been categorized as ELL by district and reported their home language was Spanish, Somali, or Hmong and also completed the Minnesota Comprehensive Assessment (MCA). Male students made up 52% of the sample and 46% were Spanish, 44% Hmong, and 10% Somali.

Invitations to Literacy, which was published by Houghton Mifflin in 1999, was the CBM-ORF administered to students in September as a part of the school's progress monitoring data collection. The MCA is a measure of reading proficiency provided by the Minnesota Department of Education. There are two types of scores that come from the MCA, a level score and a standard score. The level score ranges from 1-5. A student scoring at level 1 is described as having gaps in reading knowledge and skills. Those at a level 3 are on grade level. Those at a level 5 are considered to be above grade level. The MCA was administered in late April. Student performance is considered proficient at Level 3, or a corresponding standard score of 1420.

Given the proficiency cut off of 1420, approximately 74% of the students in the sample did not reach proficiency level. Results from the regression analysis indicate the fall CBM measure was a significant predictor of the spring MCA reading score ($F(1,1203) = 749.79$; $p < 0.001$; $r^2 = .39$). For every single word increase in CBM scores in the fall, there was an expected increase in MCA reading scores of about three points.

Thus, one would expect a CBM score of 111 words per minute to produce a state test score of 1420 (i.e., proficient). Logistic regression indicated CBM was a significant predictor of proficiency on the MCA reading test ($\chi^2 = 285.833$; $p < 0.001$) and accounted for 30% of the variance (Nagelkerke's $r^2 = .297$).

September reading CBM scores accurately classified spring state test performance about 75% of the time. It was concluded that R-CBMs are a valid tool for the purposes of screening and progress monitoring ELL students and can provide a framework of what to expect from these students. Results of this study also found that R-CBM was more accurate at determining which students would not pass the reading proficiency level test, than determining which students would actually pass. This means that R-CBM is a good indicator of later status of failing to meet the proficiency level in reading on the MCA for students from a Spanish, Hmong, and Somali population.

Richardson, Hawken, and Kircher (2012) looked at the use of CBM for ELLs. Specifically, they researched the predictive value of CBM Maze (Maze) on high-stakes performance tests for Hispanic and Spanish speaking students. The purpose of this study was to determine if Maze scores demonstrated any statistical biases toward Hispanic ethnicity or Spanish-speaking students when it comes to predicting their performance on high-stakes tests. This study was composed of fourth through sixth graders at six elementary schools in an urban school district ($N = 719$). Of these students in the sample, 26% had limited English proficiency, and 47% were non-English speakers based on the Oral Language subtest of the IDEA Proficiency Test (IPT).

The Maze test from AIMSweb was used to screen reading comprehension, and the English Language Arts Criterion Referenced Test (ELA-CRT) was used to measures

students' reading, writing and listening ability (Richardson et al., 2012). The ELA-CRT was developed by the state of Utah to assess students' ability with literacy and is given to students 2nd through 11th grade and administered in May. The AIMSweb Maze was administered to the students in their classrooms during the same month.

Researchers created three general categories for the sample: White, Hispanic, and neither White nor Hispanic (NWNH) (Richardson et al., 2012). Primary languages spoken in the home were also split into categories: English, Spanish, and neither English nor Spanish (NENS). When running analyses of bias for ethnicity, there were mean differences for both Hispanic and NWNH with lower mean scores than white students $t(706) = -3.926$, $p < .001$, and $t(706) = -4.584$, $p < .001$. For the Hispanic ethnicity, a 3.0 drop was predicted in ELA-CRT standard scores. There were no significant interactions between Maze and ethnicity for Hispanic, $t(706) = 0.039$, $p = .97$, or the NWNH group, $t(2.287) = 2.287$, $p = .07$. These results indicate that Maze can function across various ethnic groups. For those who were NWNH, there were significant differences between NENS and Maze performance $\chi 2(5) = 21.1$, $p = .001$. This means that even though there were score differences between schools, there was no difference between those who had English as their home language and the NENS students.

Overall, CBM Maze yielded low rates of false negatives (2.8%) on the ELA-CRT. When the Maze indicated the student was likely to pass the ELA-CRT, they typically did. Specifically, across demographic groups, Maze had very few false negatives (1.2-4.4%). The false positives for Maze ranged from 15 to 31% and the greatest percentage of false positives came from the students with Hispanic ethnicity or spoke Spanish at home. Maze

was found to be a strong predictor of outcomes on state testing, but should be used with caution with ELL students.

Kim, Vanderwood, and Lee (2016) examined the predictive validity of two reading CBMs and state test performance with Spanish-speaking ELLs. Third grade students (N = 522) from Southern California were administered two CBMs of reading: the DIBELS Oral Reading Fluency (R-CBM), a measure of reading fluency, and DIBELS Maze (Maze), a measure of reading comprehension. Students came from predominately lower socioeconomic families and all ELL students received regular instruction in English and 30 minutes of ELL development instruction per day. Students were grouped based on language status results from the California English Language Development Test (CELDT; Kim et al., 2016). The groups were beginning/early intermediate (B/EI; N = 291), intermediate (Int.; N = 291), and early advanced/advanced (EA/A; N = 109).

Students were screened using R-CBM and Maze in September. The R-CBM was administered individually and Maze as a group. All students were administered the California Standards Test- English Language Arts (CST-ELA) as a whole-class during the following spring. The CST-ELA is the end-of-year state assessment used to evaluate student achievement in California.

Findings revealed a strong correlation between R-CBM and CST-ELA (r = .54) and a moderate correlation between Maze And CST-ELA (r = .34). When looking more closely at correlations for different English proficiency levels, the relationship between R-CBM and CST-ELA was strong (r = .59) for the B/EI group. The two measures, Maze and CST-ELA were correlated for the intermediate group (r = .31). For the EA/A group, the relationship with R-CBM and CST-ELA was moderate (r = .36). With this being said,

the correlation between Maze and CST-ELA was not significant ($r = .15$, $p > .01$). Additional predictive validity analyses were assessed looking at the differences in predictive validity for varying ELL levels. For EA/A students, the analysis revealed that R-CBM was a significant predictor of spring CST-ELA. The B/EI and Int groups were also significant indicating there are differences in CST-ELA performances due to English proficiency levels.

Researchers also looked at the predictive accuracy of students with varying ELL levels. These analyses found that using the R-CBM cut scores, 69% students in the B/EI group who did not reach proficient reading outcomes on the spring CST-ELA were correctly identified. The other 31% of students who were not proficient on the CST-ELA were not correctly identified as at risk by R-CBM. The Maze cut-score was able to identify a higher percentage of B/EI students who were not proficient in the spring CST-ELA. It is important to note, however, that none of the students in the B/EI group scored at or above expectations on CST-ELA. For the EA/A group, the R-CBM correctly identified 26% of students who were not proficient in the spring CST-CLA.

Results of this study indicate there was a significant difference in the level of performance on CST-ELA for ELLs of varying English proficiency, however, the interaction between R-CBM and ELL level was nonsignificant. This suggests there is no difference in the predictive ability R-CBM has on CST-ELA performance based on English proficiency. This was the same result for Maze, as well. These results suggest that R-CBM is able to predict performance on CST-ELA with ELLs of all English proficiency levels. Ultimately, for Spanish speaking ELLs, Maze may not be as valuable of a measure in predicting CST-ELA scores or other standardized test scores.

Written Expression

Wallace, Campbell, Lembke, Long, and Ticha (2008) examined the use of Written Expression-CBM (WE-CBM) and predicting the success of high school students on state testing. The sample included 183 high school students with 38 of those being ELLs. Tenth grade students were asked to write for ten minutes, and their progress was marked at 3, 5, and 7 minutes. Their writing samples were scored based on words written (WW), works written correctly (WWC), correct word sequences (CWS), and correct minus incorrect word sequences (CIWS). Scores were also calculated for 3, 5, 7, and 10-minute samples of writing performance. The outcome measure used was the Minnesota Basic Standards Test/Minnesota Comprehensive Assessment (MBST/MCA). The MBST is a high stakes test required for graduation. MCA consists of high stakes tests designed to rank student performance across a continuum. The MBST/MCA writing tests are untimed and administered annually to 10^{th} graders in January. Students complete one writing sample from each. In the fall, students completed two CBM writing samples on the same day.

Results indicated students wrote steadily during the ten-minute interval and averaged about 17 words per minute. Focusing specifically on the ELL students within the sample, the ELL students wrote fewer words and less accurately than the non-ELL students. The MBST/MCA for ELL (2.65, SD = .65) and non-ELL (3.25, SD = .59) students shared differences that were statistically significant (t = 5.46, p < .001). Overall, results for both ELL and non-ELL students shared a similar pattern with the various scores calculated. Predictive validity was stronger for the ELL students than the non-ELL students. This means that using WE-CBM can be a predictive measure of ELL student's

performance on state testing. There were significant correlations found between the MBST and the WE-CBM with all scoring methods used ($p < .05$). However, there were stronger correlations with CWS and correct minus incorrect word sequences ($p < .01$).

Campbell, Espin, and McMaster (2013) examined the validity and reliability of CBMs in writing for 36 ELLs in grades 10-12 with moderate to high English language proficiency (15 males, 21 females). Of this sample, 92% spoke an African language. WE-CBM was administered for 3, 5, and 7 minutes. All other standardized administration and scoring procedures were followed. Each WE-CBM was scored based on words written (WW), correct words (CW), percent correct words (%CW), correct word sequence (CWS), and percent correct word sequence (%CWS).

Included criterion variables were the Test of Written Language-3rd Edition (TOWL-III), the Test of Emerging Academic English (TEAE), and the Minnesota Basic Skills Test (MBST), and teacher ratings of writing. The MBST is an untimed, high-stakes test which reflects students writing literacy skills and was administered in April.

Results indicated the students' pace of writing was slightly faster for narrative and expository prompts than for the picture prompts. The most reliable WE-CBM concluded that the narrative prompts resulted in the greatest number of reliability coefficients that were at or above .70 and .80. The validity of the writing tasks was investigated between the WE-CBM and the other criterion measures. On the MBST, the mean participant score was a 1.87, and a 3 is considered proficient. Overall, technical adequacy of the WE-CBM is a useful to determine ELL performance. For the MBST, there were significant correlations across all three types of WE-CBMs and across all time limits with %CW and

CIWS ($p < .01$). For the TOWL-III, there were significant correlations for the narrative and expository WE-CBM across all time limits with CW ($p < .01$).

Keller-Marguilis, Payan, Jaspers, and Brewton (2016) looked at the accuracy of WE-CBMs for students with diverse language backgrounds. The purpose of this study was to look at the validity and accuracy of WE-CBM as a predictor of performance on statewide writing achievement tests. The sample included 139 fourth-grade students from two south-western elementary schools were used for this study. Of the total sample, 19 students were identified as ELL students. There were also 31 students who were monitored who were exited from ELL in the last year or two. All of the students within the study completed WE-CBMs in English. First, the WE-CBM was administered numerous times during the fall, winter, and spring. Then, the State of Texas Assessments of Academic Readiness (STAAR) writing test was given across two days at the end of March. The STAAR statewide achievement test in writing was the mandated writing statewide test in Texas.

Results indicated the correlation between WE-CBM and the STAAR writing test were low to moderate for native English speaking students (Keller-Marguilis et al., 2016). CWS was the only indicator significantly related to STAAR performance. Overall, the highest levels of diagnostic accuracy were for the winter correct-minus-incorrect writing sequences (CIWS) score. However, there were no cut point with CIWS for ELLS that was adequate. Due to the limited sample of the ELL students within this study it is important to interpret this data with caution because it resulted in lower levels of power (Keller-Marguilis et al., 2016). Overall, for the ELL students there was a strong correlation for %CWS and CIWS winter-time points. The statistical tests also indicated

that WE-CBM indicators were not a significant predictor of statewide test performance for ELLs.

Discussion

Results from the systematic review yielded eight articles examining the use of CBMs for screening with ELLs. Specifically, five articles were found that looked at CBMs in the area of reading fluency and comprehension, and three articles looked at WE-CBMs with ELL students and the predictive validity it has on high stakes testing. It is noteworthy that across almost 30 years of literature, only eight articles were identified which examine the use of CBMs as a screener for ELLs. This review highlights the lack of research in this area.

In sum, the use of CBMs in both reading and writing varied considerably throughout the research to have predictive validity on how ELL students will perform on state tests. With this being said, results fluctuate in strength. For example, CBM scores tended to be strong predictors of state test performance for ELLs (Dominguez de Ramirez & Shapiro, 2006; Kim et al., 2016; Muyksens et al., 2009; Vanderwood et al., 2008), but may under predict performance (Vanderwood et al., 2008) or demonstrate insufficient levels of sensitivity of specificity (Kim et al., 2016; Muyskens et al., 2009). One study found that CBMs functioned differently for different language proficiency levels (Keller-Margulis et al., 2016) while another one found that predictive validity did not differ across proficiency (Kim et al., 2016).

Results suggest that measures of oral reading fluency such as R-CBM or R-CBM may be more appropriate for use with ELLs than measures of reading comprehension (Kim et al., 2016, Richardson et al., 2012). This may be due to language development and students could be expected to develop reading fluency in their non-native language earlier than comprehension.

In writing, research showed specific scoring methods within WE-CBM were valid predictor of performance on state testing (Campbell et al., 2013; Wallace et al., 2008; Keller-Marguilis et al., 2016). Specifically, %CWS, CIWS, CWS, and %WSC were sensitive to the ELL student performance on state testing. There was also research that indicated across longer time periods, there were stronger correlations with performance on state testing (Campbell et al., 2013; Wallace et al., 2008).

Despite the preliminary and promising results presented here, the research on the use of CBMs with ELLs to predict state tests is minimal compared to similar work with native English speakers (e.g., Buck & Torgesen, 2003; Grapin et al., 2017; Silberglitt & Hintze, 2005; Wood, 2006; Yeo, 2010, to list a few).

Implications

Many districts use CBMs to screen students for academic risk. As a result of screening results, students should then be categorized into groups to receive appropriate interventions. The results of this specialist project tentatively indicate CBMs are able to be used with ELL students and are not limited to only Spanish-speaking ELL students. More specifically, initial research shows that the use of CBMs in reading and writing are valid predictors of student performance on high stakes achievement testing, but should be interpreted with caution because there are varying results for both reading and writing. School personnel should continue to make decisions, particularly high-stakes decisions through the use of multiple data methods, and when possible, data appropriate for ELLs. This could be valuable information for teachers, administration, test creators, and others to determine which students are struggling academically. For R-CBM its important to know that there are varying levels of predictive validity on state tests, but some research

studies have shown that it is a useful measure with ELLs. Specifically for WE-CBMs, it is valuable to note that certain methods of scoring are more predictive of how ELL students will perform on state testing.

Limitations

This specialist project was a systematic literature review. While this format helps to provide a synthesis of the literature for readers, it is not an empirical study or meta analysis and no statistical inferences can be drawn. Also, this review focused on the use of CBM in screening and the predictive validity for high stakes testing. This review leaves out any articles which examine the use of CBMs for ELLs more broadly or for progress monitoring purposes. These types of studies were not included but would provide additional evidence about the validity of CBMs with ELLs.

Future Directions

Across academic areas, the greatest number of articles were found within reading. It is important for future studies to examine the use of CBMs with ELLs in other academic areas, particularly writing and math. Specifically, it would be valuable to look further into the use of R-CBM and early literacy for ELL students (Vanderwood et al., 2008). Many of the research studies synthesized in this review primarily focused on ELL students who spoke one language. For future studies, researchers should investigate ELL students who speak many different languages and not limited to spanish-speaking ELLs (Vanderwood et al., 2008) and languages other than Spanish. Additionally, research should look into whether findings about the use of CBMs with ELLs and the predictive validity on high stakes tests are generalizable across other criterion measures. In

conclusion, this review higlights the need for future research in this area given the significant lack of articles published in the past 29 years.

References

Buck, J., & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test* (Technical Report 1). Tallahassee: Florida Center for Reading Research.

Campbell, H., Espin, C., & McMaster, K. (2013). The technical adequacy of curriculum-based writing measures with English learners. *Reading & Writing, 26*, 431–452. https://doi-org.libsrv.wku.edu/10.1007/s11145-012-9375-6

Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly, 23*, 509–531.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics, 2*, 132–149.

Deno, S. L. (1985). *Exceptional children: Integrating research and teaching* (2nd Edition). San Diego, CA: Singular Pub Group.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*, 184–192.

Domínguez de Ramírez, R., & Shapiro, E. S. (2006). Curriculum-based measurement and the evaluation of reading skills of Spanish-speaking English language learners in bilingual education classrooms. *School Psychology Review, 35*, 356–369.

Grapin, S. L., Kranzler, J. H., Waldron, N., Joyce-Beaulieu, D., & Algina, J. (2017). Developing local oral reading fluency cut scores for predicting high-states test performance. *Psychology in the Schools, 43*(5), 527-535.

Hart, J., & Stebick, D. (2016). Making the invisible visible: RtI and reading comprehension. *New England Reading Association Journal, 51*, 43-56.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York, NY: The Guilford Press.

Keller-Margulis, M. A., Payan, A., Jaspers, K., & Brewton, C. (2016). Validity and diagnostic accuracy of written-expression curriculum-based measurement for students with diverse language backgrounds. *Reading and Writing Quarterly: Overcoming Learning Difficulties, 32*, 174-198.

Kim, J. S., Vanderwood, M. L., & Lee, C. Y. (2016). Predictive validity of curriculum-based measures for English learners at varying English proficiency levels. *Educational Assessment, 21,* 1–18.

McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., Barmer, A., Forrest Cataldi, E., and Bullock Mann, F. (2018). *The Condition of Education 2018* (NCES 2018-144). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

McMaster, K. L., Shin, J., Espin, C. A., Jung, P. G., Wayman, M. M., & Deno, S. L. (2017). Monitoring elementary students' writing progress using curriculum-based measures: Grade and gender differences. *Reading and Writing: An Interdisciplinary Journal, 30*, 2069-2091.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The Prisma Group. (2009). Preferred reporting items for systematic reviews and metal-analyses: the PRISMA statement. *Open Medicine, 3*, 123-130.

Mohr, K. J., & Mohr, E. S. (2007, February). Extending English language learners' classroom interactions using the response protocol. *The Reading Teacher, 60*, 440-450.

Muyskens, P., Betts, J., Lau, M., & Marston, D. (2009). Predictive validity of curriculum-based measures in the reading assessment of students who are English language learners. *California Association of School Psychologists*, *14*, 11-21.

National Center for Education Statistics. (2018). *English Language Learners in Public Schools*. Institute of Education Sciences, US Department of Education, Washington, D.C. Retrieved from https://nces.ed.gov/programs/coe/indicator_cgf.asp

National Center on Response to Intervention. (2010). *Essential components of RTI—a closer look at Response to Intervention*. Washington, DC. National Center on Response to Intervention. Retrieved from https://rti4success.org/sites/default/files/rtiessentialcomponents_042710.pdf

Nese, J., Park, B., Alonzo, J., & Tindal, G. (2011). Applied curriculum-based measurement as a predictor of high-stakes assessment: Implications for researchers and teachers. *The Elementary School Journal*, *111*, 608-624.

Reed, R. (2013). English-language learners. In C. E. Cortés (Ed.), *Multicultural America: A multimedia encyclopedia* (Vol. 1, pp. 792-793). Thousand Oaks, CA. SAGE Publications, Inc. doi: 10.4135/9781452276274.n294

Rhodes, R. L., Ochoa, S. H., & Oritz, S. O. (2005). *Assessing culturally and linguistically diverse students: A practical guide.* New York, NY: The Guilford Press.

Richardson, R., Hawken, L., & Kircher, J. (2012). Bias using Maze to predict high-stakes test performance among Hispanic and Spanish-speaking students. *Assessment for Effective Intervention, 37,* 159–170. https://doi.org/10.1177/1534508411430320.

Robertson, K., & Ford, K. (2009). *Language acquisition: An overview.* Retrieved from http://www.colorincolorado.org/article/language-acquisition-overview

Silberglitt, B., & Hintze, J. (2005). Formative assessment using CBM-R cut scores to track progress toward success on state-mandated achievement tests: A comparison of methods. *Journal of Psychoeducational Assessment, 23,* 304–325.

Texas Education Agency. (1997). *TAAS technical digest.* Austin, TX: Author.

Unruh, S., & McKellar, N. (2017). *Assessment and intervention for English language learners translating research into practice*. https://doi.org/10.1007/978-3-319-52645-4

VanDerHeyden, A. M., & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention, 30,* 5-32.

Vanderwood, M. L., Linklater, D., & Healy, K. (2008). Predictive accuracy of nonsense word fluency for English language learners. *School Psychology Review, 37,* 5–17.

Walker, A., Shafer, J., & Iiams, M. (2004). "Not in my classroom": Teacher attitudes towards English language learners in the mainstream classroom. *NABE Journal of Research and Practice, 2,* 130-160.

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

Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment, 11,* 85–104.

Yeo, S. (2010). Predicting performance on state achievement tests using curriculum-based measurement: A multilevel meta-analysis. *Remedial and Special Education 31*, 412-422. doi: 10.1177/0741932508327463.