eJEP
eJournal of Education Policy

Using existing data to predict reading achievement among English language learners in the United States

To cite this article: Adams, J. L., Persinger, L. L, Bohan, K. J., & Horn, R. A. (2020). Using existing data to predict reading achievement among English language learners in the United States. eJournal of Education Policy, 21(1). https://doi.org/10.37803/ejepS2008

John L. Adams, Ph.D.
Assistant Professor of Psychology
Olivet Nazarene University, College of Arts and Sciences
1 University Ave.
Bourbonnais, IL 60914
Jadams3@olivet.edu

Lisa L. Persinger, Ph.D.
Assistant Professor of Educational Psychology
Northern Arizona University, College of Education
801 S. Knoles Ave.
Flagstaff, AZ 86011-5774
Lisa.Persinger@nau.edu

Kathy J. Bohan, Ed.D.
Associate Professor of Educational Psychology
Northern Arizona University, College of Education
801 S. Knoles Ave.
Flagstaff, AZ 86011-5774
Kathy.Bohan@nau.edu

Robert A. Horn, Ph.D.
Associate Professor of Educational Psychology
Northern Arizona University, College of Education
801 S. Knoles Ave.
Flagstaff, AZ 86011-5774
Robert.Horn@nau.edu

Author Note.

This research is an extension of a study completed for the first author's doctoral dissertation. Correspondence concerning this article should be addressed to John L. Adams, Department of Behavioral Sciences, Olivet Nazarene University, Bourbonnais, IL, 60914.

Abstract
This study investigates the utility of existing measures of Oral Reading Fluency (ORF), reading retell, and language proficiency for predicting reading achievement among Spanish speaking English language learners in the United States. Consistent with previous findings, ORF predicts a large proportion of the variance in reading achievement. Additionally, retell and language proficiency are significant predictors above and beyond ORF. These findings support the potential utility of using existing measures for monitoring the reading progress of students who are second language learners within a district. Implications for policymakers are discussed.

Keywords: Progress Monitoring, English Language Learners, Reading Achievement, Statewide Reading Achievement Tests, Language Proficiency

Context of the Problem

Reading is an important skill with far-reaching impact well beyond the classroom (Palani, 2012). In the U.S. state of Arizona, schools are evaluated on students' reading growth (Arizona Department of Education, 2019). To incentivize high academic standards, the state awards additional funds to schools that show exceptional performance in this area (Arizona Department of Education, 2019). Yet, some of the most at-risk students continue to fall behind (Sanders et al., 2018).

Among students at risk for reading failure, Spanish-speaking English Language Learners (ELLs) make up an increasing proportion of students learning a second language in Arizona and across the U.S. (English Language Learners in Public Schools, 2018). Spanish-speaking students face a different set of challenges than monolingual English-speaking students and underperform in reading relative to their ELL peers (Roberts, Mohammed, & Vaughn, 2010). To investigate the impact of linguistic diversity on reading skills' development, we first turn to the role of language in reading acquisition.

The Simple View of Reading

According to the Simple View of Reading , reading skills are acquired through a combination of language and decoding skills (Gough & Tunmer, 1986; Hoover & Gough, 1990; McCardle, Scarborough, & Catts, 2001). The Simple View of Reading asserts that reading comprehension is a product of oral language skills and word level reading skills. Consistent with this model, numerous studies have shown that language differences contribute to differences in reading performance among ELLs (e.g., Geva & Farnia, 2012; Proctor, Carlo, August, & Snow, 2005; Uchikoshi, Yang, & Liu, 2018). Jeon and Yamashita (2014) conducted a meta-analysis of reading comprehension studies, finding vocabulary knowledge, grammar knowledge, and decoding skills were the strongest correlates of second-language reading skills among second language learners.

Vocabulary and listening comprehension, in particular, are strong predictors of reading comprehension and become more important following the mastery of basic reading skills (Adlof, Catts & Lee, 2010; Babayiğit, 2014; Goodwin, August, & Calderón, 2015; Gottardo, Mirza, Koh, Ferreira, & Javier, 2018; Lindsey, Manis, & Bailey, 2003; Vaughn et al., 2019). Evidence suggests language differences contribute to continued disparities in comprehension skills after decoding gaps between ELL students and their non-ELL peers close (Droop & Verhoeven, 2003; Vaughn et al., 2019). The correlation between language skills and reading comprehension has been shown to be even stronger among ELL students than for non-ELL students (Babayiğit, 2014; Proctor, Carlo, August, & Snow, 2005).

With the importance of language differences established, we are left with the practical question of how to address these differences. The Multitiered Systems of Support (MTSS) framework provides a structure for addressing student needs in an inclusive manner. MTSS is a structured approach providing students with interventions at the required intensity based on frequent use of meaningful data. MTSS follows a tiered structure in which higher tiers call for more intensive academic interventions in small, homogenous groups (Ball & Christ, 2012). Within this framework, progress monitoring with Curriculum Based Measures (CBMs) is one approach for

identifying and addressing students' needs (Deno, 1985; 2003; 2016). Understanding CBMs is critical for making meaning of outcomes studies using them.

Curriculum Based Measures and Reading Outcomes

CBMs are brief probes that test students' grade-level academic skills (Deno, 1985). They are effective at identifying and monitoring students' academic needs (Deno, 2003). These measures are widely used in U.S. schools in the early grades.

Oral reading fluency (ORF) predictive utility is strongly supported in the literature. ORF measures are text passages students read aloud for usually one minute while the examiner marks errors. Scores record the number of correct words read and the number of errors. Reschly, Busch, Betts, Deno, and Long (2009) conducted a meta-analysis of 40 correlational studies including ORF measures finding moderate to strong correlations (average of $r = .67$) between ORF and high-stakes reading assessments for students from grades 1-6. Yeo (2010) performed a follow-up meta-analysis of 27 studies employing multiple reading CBMs as performance predictors on state achievement tests, finding an average correlation of .68, in the moderately high range. Authors of both studies concluded that using reading CBMs to predict reading outcomes is strongly supported (Reschly et al., 2009; Yeo, 2010). Additionally, Kilgus, Methe, Maggin, and Tomasula (2014) conducted a meta-analysis of the diagnostic accuracy of ORF, finding support for the sensitivity and specificity of ORF cut scores for predicting overall reading achievement on high stakes tests. The exact cut scores used to achieve desired sensitivity and specificity varied between studies, but the general finding of predictive utility was upheld.

Fewer studies support the predictive utility of other (non-ORF) reading CBMs (Fuchs, Fuchs & Compton, 2004), but comprehension CBMs have shown promise. Maze is a reading comprehension CBM where students read a passage wherein three word options complete a sentence at about every seventh word. Student scores are based on the number of words correctly restored in the passage. Maze measures have been shown to be a significant predictor of reading achievement, but not as strong a predictor as ORF (Wiley & Deno, 2005).

Beyond Maze CBMs, Shapiro, Fritschmann, Thomas, Hughes, and McDougal (2014) found that a 10-point retell quality rating predicted a small portion of variance above and beyond ORF in a sample of third grade students. Retell measures typically involve asking students to retell the story read following completion of an ORF probe. Scores can be based on overall quality of their retell or on the number of words used to retell the story. There is promise in considering the additional predictive value comprehension CBMs have in identifying how children will perform on high stakes reading tests.

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Retell measure, one type of commercially available CBM, includes a similar quality rating measure of passage comprehension as used by Shapiro et al. (2014), in addition to the number of words used in a correct retelling. Missall, Hosp and Hosp (2019) found the DIBELS Retell subtest was significantly correlated to standardized tests of reading achievement, but Retell and other measures showed less correlation to performance than did ORF. However, Missall et al. (2019) only examined the number of words retold and did not include consideration of the retell quality ratings. Our study includes a comprehension quality rating as one measure investigated.

For MTSS to be most effective, assessments must be used appropriately and provide information necessary to identify students needing intervention as well as directing those interventions (Deno, 2016). Yeo (2009) concluded that language proficiency was a potentially important variable, but insufficient evidence to substantiate use of language proficiency measures existed at that time. Few studies have looked specifically at the predictive utility of CBMs for ELL students. The next section will summarize findings of the few predictive utility studies.

CBMs and Prediction of Reading Outcomes for ELLs

Scheffel, Lefly, and Houser (2016) found the predictive utility of DIBELS ORF (DORF) was roughly equivalent for all students. Notably, their analyses did not account for differing levels of language proficiency among ELL students. Kim, Vanderwood, and Lee (2016) found no significant difference in DORF and DIBELS Daze predictive accuracy across language proficiency levels. Researchers investigated comparisons between students falling in different language proficiency categories, but did not account for oral listening and speaking development separately from reading and writing. Similarly, Burns et al. (2017) employed categorical language proficiency data, using overall English language proficiency level to predict student growth on an oral reading fluency CBM. They found students at the lowest levels of proficiency showed the greatest gains in words read correctly in second and third grades. The study also found that language proficiency level predicted a small proportion of variance in a computerized measure of reading achievement. This provided the opportunity to better handle language proficiency level complexity but did not deal with complexity associated with language proficiency with listening versus speaking versus reading versus writing. The present study builds on this finding by employing scaled scores from a state-wide measure of language proficiency broken down to listening and speaking, thus removing the proficiency subtests based on students' English reading and writing skills. This provides a cleaner measure of language proficiency that is not influenced by reading and writing skill levels.

There is a clear need for investigating language proficiency's potential contribution to CBMs' predictive accuracy. Existing studies (e.g., Kim, Vanderwood, & Lee, 2016; Scheffel, Lefly, & Houser, 2016) have either ignored proficiency level or employed it as a unitary construct without considering that oral language develops before reading and writing (Shanahan, 2006).

Studies reviewed here indicate progress monitoring data and language proficiency information, specifically listening and speaking skills, have potential for predicting ELL students' performance in reading beyond CBMs. Though CBMs are effective, they are best used in combination with other sources of information (Deno, 2016). Improving screening procedures' accuracy is imperative for prudent use of school resources.

Current Study

This study aims to improve the predictive model of ELL students' performance on high stakes reading tests by extending prior research with additional relevant variables. We investigate the predictive utility of reading comprehension measures and language proficiency above and beyond ORF. Previous studies have provided strong support for ORF utility and some support for the utility of comprehension CBMs and language proficiency. We seek to build on these findings by employing a previously underexplored measure of reading comprehension quality

and a language proficiency composite score that eliminates reading and writing subscores, thus providing a cleaner measure of oral language proficiency status in the model.

Research Questions

Research questions for the current study are:

1. To what extent do ORF and reading comprehension CBMs predict reading performance among ELLs as measured by state high stakes reading tests?
2. To what extent might oral language proficiency add to the predictive utility of the CBMs among ELLs?

Method

Sample

From an initial sample of 2,865 students drawn for a broader study, the current study sample included 231 Spanish-speaking ELL students. Data were gathered from a large school district in the southwestern United States from the 2014-2015 and 2015-2016 school years. Students were included in the data set if they were identified as ELLs by the Primary Home Language Other Than English (PHLOTE) form used in Arizona public schools. Students were removed from the data set if they were missing one or more of the scores used in the predictive model. One hundred percent of the students included in the sample were Hispanic. Fifty-seven percent were male, 43% were female. AzMERIT scores include four categories from Minimally Proficient to Highly Proficient. Eighty percent scored in the Minimally Proficient range, 11% scored in the Partially Proficient range, and 9% scored in the Proficient range. None of the students in our sample scored in the Highly Proficient range.

Instruments

Arizona English Language Learner Assessment (AZELLA). The Arizona English Language Learner Assessment (AZELLA) is Arizona's language proficiency test used state-wide to determine if students require services due to different linguistic performance. The test measures Listening, Speaking, Reading, and Writing proficiency, providing a Total score. In the present study, standard scores from the Listening and Speaking subdomains were summed. This researcher-created composite was used to represent language ability unaffected by reading and writing skills. This composite is one of the predictor variables for the study. These scores were taken from the students' second grade year in spring 2015.

***Reliability and validity*.** For the present study, AZELLA Listening and Speaking subtests' reliability are most relevant. Test authors report a Cronbach's alpha of .66 for second grade on the Listening subtest. Test authors report a Cronbach's alpha of .81 for the second grade Speaking subtest. Authors report a panel of educational experts was used to evaluate content and perform field-testing.

Dynamic Indicators of Basic Early Literacy Skills Next Edition (DIBELS Next). DIBELS Next is a battery of CBMs used to assess basic reading skills development, primarily through one-minute reading tasks (Good et al., 2013). Measures of interest for the present study are those administered to second grade students in the Spring semester: DIBELS Oral Reading

Fluency-Words Read Correctly (DORF WRC), Recall, and Retell Quality. DORF WRC tests students' ability to read a grade-level passage aloud for one minute. The student's score is equal to the number of words read accurately within the time limit.

Retell Quality is then rated on a scale of one to four, where four represents the highest quality Retell. A rating of one indicates two details retold from the reading passage. A two indicates the student successfully recalled three or more details. A rating of three indicates three or more details retold in appropriate sequence. A four rating indicates the student successfully retold three or more details in sequence and captured the main idea of the passage. These, Retell, Retell Quality and DORF WRC scores were also collected during students' second grade year in the spring of 2015 and are additional predictor variables for this study.

Reliability and validity. Reliability coefficients for most measures fall well above the standard of .70 (Cortina, 1993) for second and third grade students for the Total score. However, the test-retest reliability coefficient for Retell is reported at .27 for second grade students. The test-retest reliability coefficient for Accuracy is .57 for second grade students.

The test makers treat Retell Quality scores as a categorical score; they are not reported as interval variables. However, there is a precedent in the literature for treating Likert scale variables as scale variables when characteristics are similar (Norman, 2010). Findings will be subjected to tests of validity as a part of the present study.

Arizona's Measurement of Educational Readiness to Inform Teaching (AzMERIT)
AzMERIT is a computer-administered achievement test for Arizona administered annually starting in third grade. The measure contains both fixed response (i.e., multiple choice, multiple response, matching) and open-ended response items (American Institutes for Research, 2017) and is used to determine overall reading skill proficiency for public school students. Every state in the United States has a similar test. Students' AzMERIT scores were taken from their third-grade year in spring 2016.

Reliability and validity. AzMERIT internal consistency coefficients were reported through Cronbach's alpha for each grade level and major subgroup tested. For grade 3, Cronbach's alpha value was .90, well above the commonly recommended standard of .70 (Cortina, 1993). For Hispanic third grade students, the value was similar at .88.

The AzMERIT technical report states many of the items were initially created for Utah's state achievement test. The report details development phases and validation for the measure, which consisted mostly of context experts' item review for alignment with grade-appropriate standards (American Institutes for Research, 2017).

Data Screening and Analysis

Table 1
Correlations Matrix for Predictor Variables and AzMERIT ELA

Variable	1	2	3	4	5
1. AZMERIT	-				
2. DORF	.58**	-			
3. Retell	.44**	.58**	-		
4. Quality	.44**	.44**	.64**	-	
6. Listening/Speaking	.41**	.35**	.31**	.34**	-

* $p < .05$ (two-tailed). ** $p < .01$ (two-tailed).

Participants were screened from the larger sample to include only Spanish-speaking ELL students. Students were included in the final data set only if Spanish was indicated as their primary language, language most spoken, and language first acquired. Students with missing scores on any of the predictor or criterion variables were excluded. The final sample consisted of 231 students.

The data were tested to ensure assumptions of linear regression were met. Data were screened for missing, normality, and homoscedasticity of error terms. One of the variables, Retell Quality, demonstrated a positive univariate skew that was not correctable by standard transformation methods. As a result, this variable was retained in the original format for primary analyses. Correlations and scatter plots indicated a linear relationship between the predictor variables and the AzMERIT. Examination of Q-Q and Residual plots indicated assumptions of homoscedasticity and multivariate normality were met. Variance Inflation Factors fell below 10 and Tolerance values fell above .4. All multivariate assumptions for primarily analyses were met.

Hierarchical regression analysis was run in three steps to address both research questions. In the first step, only DORF WRC was entered. At step two, additional DIBELS Next variables (Accuracy, Retell, Retell Quality) were included. In the third and final step, the Listening/Speaking variable (AZELLA) was added.

Results

Research Question 1

To address research question one, researchers ran a prediction model using ORF (DORF WRC) and reading comprehension CBMs (Retell and Retell Quality) as predictors for the state reading test (AzMERIT) in steps one and two of the hierarchical regression model. Step 1 contained only the DORF WRC variable for spring 2015 as the predictor for AzMERIT spring 2016. We ran the model with only this variable to sort out the improvement in prediction with additional CBMs and because ORF is the best supported and understood CBM in the literature. As expected and consistent with prior research, DORF WRC was a significant predictor of AzMERIT score, F (1, 234) = 122, $R2$ =.34, $p < .001$ at step one. These findings indicate that DORF WRC accounts for 34% of variance in children's AzMERIT test scores in English Language Arts.

In the second step, Retell, and Retell Quality were added to see what additional variance in test performance might be explained beyond the DORF WRC. In the second phase, only DORF WRC and Retell Quality were significant predictors, F (4, 231) = 62.66, $R2$ Change = .04, $p < .001$ where the addition of Retell Quality improved the prediction model by 4%.

Research Question 2

Research question two focused on whether ELL students' oral language proficiency improves the prediction for third grade state test performance. To address this question the AZELLA Listening/Speaking Composite was entered in step three to see if language proficiency improved the prediction of high-stakes reading test performance. The Listening/Speaking Composite contributed significantly to the prediction of variance in AzMERIT above and beyond the variables in the first two steps, adding an additional 4% to the variance explained, F (5, 230) = 56.88, $R2$ Change = .04, $p < .001$. Table 2 provides a summary of these models.

Table 2.
Summary of Hierarchical Regression Analysis (N = 231)

Variable/Model	β	t	Sig.	F	R^2	Sig.
Model 1				122.3	.34	< .001
DORF WRC	.59	11.1	< .001			
Model 2				48.4	.39	< .001
DORF WRC	.48	7.53	< .001			
Retell	.01	.186	.853			
Quality	.22	3.25	.001			
Model 3				40.9	.42	< .001
DORF WRC	.43	6.87	< .001			
Retell	.01	.076	.939			
Quality	.18	2.74	.007			
Listening/Speaking	.19	3.41	< .001			

R^2 Change (Model 2) = .04 R^2 Change (Model 3) = .03

Discussion

Consistent with previous studies, the present study supports ORF's utility as a predictor for children's performance on high-stakes state reading tests. ORF alone predicted over a third of variance in test performance for ELL students. Our study extended prior research by investigating the contribution of Retell and Retell Quality CBMs above and beyond ORF in this prediction. Additionally, we added a unique oral language proficiency score to determine its value in the prediction model.

For ELL students, Retell Quality was a significant predictor above and beyond ORF for the second and third steps of the regression model. The Listening/Speaking composite derived from the AZELLA was a significant contributor in the step three. The amount of additional explained

variance was small relative to ORF; however, these results provide support for the importance of attending to language proficiency and suggest a possible avenue for future development in progress monitoring second language learners.

Given the final model's statistical significance, we now turn to the model's individual components with attention to features that may be useful to policy makers. We discuss the predictive validity of Retell Quality and the Listening/Speaking composite as continuous variables. These variables provide additional information when used together to get a better understanding of ELL students' progress toward passing high stakes tests.

Retell Quality

Given the limited range of scores, the contribution of Retell Quality above and beyond ORF is surprising. In fact, the measure performed similarly to the 10-point measure reported by Shapiro, Fritschmann, Thomas, Hughes, and McDougal (2014). This is striking for two reasons. First, the range of DIBELS Retell is restricted relative to the measure used in the Shapiro et al. study. Range restriction typically reduces correlation, which in turn, reduces the variable's impact in the model. Second, data used in our study were gathered by teachers and staff in the course of regular educational service delivery. In contrast, Shapiro et al. (2014) gathered data with help from trained doctoral students. This difference in data collection methods speaks to the ecological validity of our findings and validates the notion that improved predictions are possible within the current education system, with current measures, and with current levels of training. This is good news for district policy makers.

Future research with similar scales should focus on potential scale and population characteristics that may influence the predictive utility of retell measures. Some notable differences in the samples used in our study reveal potential starting points for investigating factors that may affect the predictive utility of retell. For example, the sample in the Shapiro et al. (2014) study consisted primarily of Caucasian students, while our study included only Hispanic students. However, the overall correlation of Retell Quality with the state-wide reading test among English-only students in Adams (2017) was stronger than the correlation between Shapiro's (2014) measure and the state-wide reading test. This may be due to a number of factors including 1) differences in state-wide tests, 2) differences in student populations under investigation, or 3) characteristics of retell measures. Partial replication studies that include Retell Quality and similar measures with different student populations and state-wide tests help sort out these potential explanations.

Listening/Speaking Composite as a Significant Predictor

Consistent with the Simple View of Reading, Retell Quality and the Listening/Speaking composite added to the prediction of state reading test performance above and beyond ORF. Given ORF's strength as a predictor, the contribution of these variables provides strong evidence in supporting the utility of language proficiency monitoring for ELL students. To date, this is the first study to employ language proficiency scaled scores in predicting reading outcomes. Past studies made use of categorical scores. Additionally, this is the first study in which reading and writing portions of a language proficiency measure were systematically excluded from analysis. Since reading is the construct under investigation, it stands to reason that a reading subtest within

a language measure might confound results with respect to the contribution of language abilities to reading.

A Step Toward Practical Applications for Policy Makers

While the present findings cannot and should not be implemented as-is in a screening or placement procedure for tiered instruction, our study reveals the potential of existing measures to be used in this way in the future. Consistent with Deno's (2016) recommendation, we combine measures to produce greater predictive accuracy. Most importantly, the present study employs data gathered by teachers in the course of normal instruction. Given budgetary and personnel constraints, school policy makers stand to benefit from research that prioritizes practicality and ease of implementation.

Additionally, this research suggests that district leaders should ensure that retell quality measures are required at the building level. This variable and other retell variables are not often seen in the literature, which may reflect limited use or perceived value in schools. Our research suggests that this notion be revisited.

Ultimately, schools would benefit from a predictive formula that would allow educators to differentially predict second language students' performance using CBM and language proficiency data. While the current findings support the viability of such an approach, further investigation into the specific measures and the generalizability across populations is needed before we can have confidence in a predictive formula.

Limitations and Directions for Future Research

Though it adds to the potential utility of the findings, using existing data limited the present study in several ways. Item level data were unavailable which restricted analysis to total scores on each scales and subscales used. Item level data would be informative regarding specific sub-skills that contribute most to the prediction of reading outcomes. Only general conclusions about the potential utility of these data can be drawn based upon these results. Using data that are already gathered in schools provides potential for efficient improvement of data-based decision making; however, future research addressing the same questions with different measures of reading, comprehension, and language skills will help us understand whether the general model holds up.

Future research should investigate the extent to which models like ours add to the sensitivity and specificity of CBMs as predictors of reading achievement among ELL students. Additionally, research should contribute to creating decision criteria that could be put into practice in schools. Due to the relatively small sample of ELL students meeting our criteria and the disproportionate number of these students who scored in the Minimally Proficient range, this was not possible in the current study.

The evidence from this study supports an optimistic outlook for districts seeking to improve prediction of ELL students' performance on critical reading tests. We provide initial support for the viability of developing a predictive formula that would allow policy makers to differentially predict ELL students' performance given CBMs and language proficiency test results from as early as twelve months prior the state reading test. Additionally, this study demonstrates the

potential of repurposing data school systems are already gathering to glean useful information. As education policy makers look toward the future of data-driven decision making in increasingly diverse schools, both findings support the exploration of potential improvements to current monitoring practices.

References

Adams, J. L. (2017). *Employing progress monitoring measures to create homogenous small groups and predict reading performance among English language learners and non-English language learners* (Doctoral dissertation). Retrieved from ProQuest Dissertations Publishing. (Order Number 10687652)

Adlof S. M., Catts H. W., & Lee J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities, 43*, 332–345.doi: https://doi.org/10.1177/0022219410369067

American Institutes for Research (2017). *Annual technical report: Arizona statewide assessment in English language arts and math* . Report prepared for Arizona Department of Education. https://cms.azed.gov/home/GetDocumentFile?id=59c2e9ac3217e10dac1f2ffd

Arizona Department of Education (2019). *FY 2020 results -based funding distribution, eligibility requirements, and accounting and expenditure reporting* (School Finance Memorandum 20-023). https://sf.azed.gov/sf/numberedmemos/FY2020%20Numbered%20Memos/20-023.pdf

Ball, C. R., & Christ, T. J. (2012). Supporting valid decision making: Uses and misuses of assessment data within the context of RTI. *Psychology in the Schools, 49*(3), 231-244.doi: https://doi.org/10.1002/pits.21592

Babayiğit, S. (2014). The role of oral language skills in reading and listening comprehension of text: A comparison of monolingual (L1) and bilingual (L2) speakers of English language. *Journal of Research in Reading, 37*(S1), S22-S47. doi: https://doi.org/10.1111/j.1467-9817.2012.01538.x

Burns, M. K., Frederick, A., Helman, L., Pulles, S. M., McComas, J. J., & Burns, Aguilar, L. (2017). Relationship between language proficiency and growth during reading interventions. *The Journal of Educational Research, 110*(6), 581-588.doi: https://doi.org/10.1080/00220671.2016.1158689

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1),98.doi:https://doi.org/10.1037/0021-9010.78.1.98

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232. doi: https://doi.org/10.1177/001440298505200303

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*(3), 184-192. doi:https://doi.org/10.1177/00224669030370030801

Deno, S. L. (2016). Data-based decision-making. In Jimerson S., Burns M., VanDerHeyden A. *Handbook of Response to Intervention* (pp. 9-28). Springer, Boston, MA. doi:https://doi.org/10.1007/978-1-4899-7568-3_2

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first-and second-language learners. *Reading Research Quarterly, 38*(1), 78-103. doi:https://doi.org/10.1598/rrq.38.1.4

Institute for Education Sciences: National Center for Education Statistics. (2019, May). *English Language Learners in Public Schools*. Retrieved from https://nces.ed.gov/programs/coe/indicator_cgf.asp

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71*(1), 7–21. doi: https://doi.org/10.1177/001440290407100101

Geva, E., & Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of reading comprehension in ELL and EL1. *Reading and Writing, 25*, 1819-1845. doi: https://doi.org/10.1007/s11145-011-9333-8

Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K.A., & Latimer, R.J. (2013). *DIBELS Next Technical Manual.* Dynamic Measurement Group. https://dibels.org/dibelsnext

Goodwin, A. P., August, D., & Calderón, M. (2015). Reading in multiple orthographies: Differences and similarities in reading in Spanish and English for English learners. *Language Learning, 65*(3), 596-630. doi:https://doi.org/10.1111/lang.12127

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10. doi: https://doi.org/10.1177/074193258600700104

Gottardo, A., Mirza, A., Koh, P. W., Ferreira, A., & Javier, C. (2018). Unpacking listening comprehension: The role of vocabulary, morphological awareness, and syntactic knowledge in reading comprehension. *Reading and Writing, 31(*8), 1741-1764. doi: https://doi.org/10.1007/s11145-017-9736-2

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*, 127-160. doi: https://doi.org/10.1007/bf00401799

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*(1), 160-212. doi: https://doi.org/10.1111/lang.12034

Kilgus, S. P., Methe, S. A., Maggin, D. M., & Tomasula, J. L. (2014). Curriculum-based measurement of oral reading (R-CBM): A diagnostic test accuracy meta-analysis of

evidence supporting use in universal screening. *Journal of School Psychology, 52*(4), 377-405. doi: https://doi.org/10.1016/j.jsp.2014.06.002

Kim, J. S., Vanderwood, M. L., & Lee, C. Y. (2016). Predictive validity of curriculum-based measures for English learners at varying English proficiency levels. *Educational Assessment, 21*(1), 1–18. doi: https://doi.org/10.1080/10627197.2015.1127750

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology, 95*, 482-494. doi: https://doi.org/10.1037/0022-0663.95.3.482

McCardle, P., Scarborough, H. S., & Catts, H. W. (2001). Predicting, explaining, and preventing children's reading difficulties. *Learning Disabilities Research & Practice, 16*, 230-239. doi: https://doi.org/10.1111/0938-8982.00023

Missall, K. N., Hosp, M. K., & Hosp, J. L. (2019). Reading proficiency in elementary: Considering statewide testing, teacher ratings and rankings, and reading curriculum-based measurement. *School Psychology Review, 48*(3), 267–275. doi: https://doi.org/10.17105/spr-2017-0152.v48-3

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education, 15*(5), 625-632. doi: https://doi.org/10.1007/s10459-010-9222-y

Palani, K. K. (2012). Promoting reading habits and creating literate society. *Researchers World, 3*(2), 90.

Proctor, C. P., Carlo, M., August, D., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology, 97,* 246-256. doi: https://doi.org/10.1037/0022-0663.97.2.246

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47*(6), 427-469. doi: https://doi.org/10.1016/j.jsp.2009.07.001

Roberts, G., Mohammed, S. S., & Vaughn, S. (2010). Reading achievement across three language groups: Growth estimates for overall reading and reading subskills obtained with the Early Childhood Longitudinal Survey. *Journal of Educational Psychology, 102*(3), 668–686. doi: https://doi.org/10.1037/a0018983

Sanders, S. M., Durbin, J. M., Anderson, B. G., Fogarty, L. M., Giraldo-Garcia, R. J., & Voight, A. (2018). Does a rising school climate lift all boats? Differential associations of perceived climate and achievement for students with disabilities and limited English proficiency. *School Psychology International, 39*(6), 646-662. doi: https://doi.org/10.1177/0143034318810319

Scheffel, D., Lefly, D., & Houser, J. (2016). The predictive utility of DIBELS reading assessment for reading comprehension among third grade English language learners and English speaking children. *Reading Improvement, 53*(2), 87-100.

Shanahan, T. (2006). Relations among Oral Language, Reading, and Writing Development. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), Handbook of writing research (p. 171–183). The Guilford Press.

Shapiro, E. S., Fritschmann, N. S., Thomas, L. B., Hughes, C. L., & McDougal, J. (2014). Concurrent and Predictive Validity of Reading Retell as a Brief Measure of Reading Comprehension for Narrative Text. *Reading Psychology, 35*(7), 644–665. doi: https://doi.org/10.1080/02702711.2013.790328

Uchikoshi, Y., Yang, L., & Liu, S. (2018). Role of narrative skills on reading comprehension: Spanish–English and Cantonese–English dual language learners. *Reading and Writing: An Interdisciplinary Journal, 31*(2), 381–404. doi: https://doi.org/10.1007/s11145-017-9790-9

Vaughn, S., Roberts, G., Capin, P., Miciak, J., Cho, E., & Fletcher, J. M. (2019). How initial word reading and language skills affect reading comprehension outcomes for students with reading difficulties. *Exceptional Children, 85*(2), 180-196. doi: https://doi.org/10.1177/0014402918782618

Wiley, H. I., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26*(4), 207–214. doi: https://doi.org/10.1177/07419325050260040301

Yeo, S. (2010). Predicting performance on state achievement tests using curriculum-based measurement in reading: A multilevel meta-analysis. *Remedial and Special Education. 31*(6), 412-422. doi: https://doi.org/10.1177/0741932508327463