

# Aspects of Technical Adequacy of an Early-Writing Measure for English Language Learners in Grades 1 to 3

Assessment for Effective Intervention  
2021, Vol. 47(1) 59–63  
© Hammill Institute on Disabilities 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1534508420947157  
aei.sagepub.com



R. Alex Smith, PhD<sup>1</sup>, and Erica S. Lembke, PhD<sup>2</sup>

## Abstract

This study examined the technical adequacy of Picture Word, a type of Writing Curriculum-Based Measurement, with 73 English learners (ELs) with beginning to intermediate English language proficiency in Grades 1, 2, and 3. The ELs in this study attended schools in one midwestern U.S. school district employing an English-only model of instruction and spoke a variety of native languages. ELs completed two forms of Picture Word in the fall, winter, and spring. The criterion measure, a common English language proficiency assessment, was administered in the winter. Results indicated that Picture Word was not appropriate for the first-grade EL participants but showed promise for second- and third-grade ELs.

## Keywords

English learners, writing, assessment

English learners (ELs) represent a significant (i.e., 10%) and diverse segment of the K–12 population (National Academies of Sciences, Engineering, and Medicine [NAEM], 2017) that are significantly underperforming academically in comparison with their native English-speaking (NES) peers (NAEM, 2017). This achievement gap is more pronounced on tasks with higher levels of language demand, such as writing (Abedi & Gándara, 2006). Writing is a key variable in overall literacy development and academic success (Graham & Hebert, 2011). Although early identification of risk and early intervention can promote long-term academic success in writing (Berninger & Amtmann, 2003; Graham & Hebert, 2011), young ELs are less likely to receive early intervention (NAEM, 2017).

## Key Variables in Screening ELs

Screening is essential for early identification of risk and access to early intervention (Salvia et al., 2017). Two key variables to consider when screening ELs for early intervention include, but are not necessarily limited to, English language proficiency (ELP) and language of instruction (Abedi & Gándara, 2006; August & Shanahan, 2006). ELP is generally reported according to a standardized assessment employing subtests in reading, writing, receptive oral language, and expressive oral language and is described as beginning, intermediate, advanced, and proficient (Fox & Fairbairn, 2011).

ELs with beginning ELP are building their receptive English language skills, may produce writing that is repetitive (e.g., employs the same sentence structure), and often benefit from vocabulary and/or sentence supports to produce written English (Roseberry-McKibbin & Brice, n.d.). ELs with intermediate ELP engage in expressive communication more readily while their writing may contain errors that are directly influenced by their native language (Roseberry-McKibbin & Brice, n.d.). ELs with advanced ELP are approaching mastery of basic oral English communication and are readily expanding their expressive English skills until they reach proficiency (generally described as being similar to a same-age NES peer). ELs may present as being fluent in English because they are able to communicate effectively using basic interpersonal oral English, but they are still learning academic vocabulary and the grammar rules of written English (Roseberry-McKibbin & Brice, n.d.). Therefore, practitioners should be careful to compare the progress of any EL with the performance of their EL peers with similar ELP (Abedi & Gándara, 2006). Abedi and Gándara (2006) suggested establishing validity

<sup>1</sup>The University of Southern Mississippi, Hattiesburg, USA

<sup>2</sup>University of Missouri, Columbia, USA

## Corresponding Author:

R. Alex Smith, School of Education, The University of Southern Mississippi, 118 College Drive, Hattiesburg, MS 39406, USA.  
Email: r.alexander.smith@usm.edu

of measures with ELs using criterion measures that have been specifically normed with ELs, such as many of the common ELP assessments.

Oral language is a critical factor in the literacy development of all children, but the English literacy development of ELs is influenced by *both* their oral English proficiency and their native language proficiency (August & Shanahan, 2006). An EL's native language serves as a resource when learning English, and research indicates that ongoing instruction in an EL's native language supports their literacy development in English (August & Shanahan, 2006). However, many ELs receive English-only instruction in U.S. schools, and research indicates that ELs receiving English-only instruction progress differently in English literacy than their peers receiving bilingual instruction (August & Shanahan, 2006). Therefore, studies should document the language of instruction EL participants receive in addition to the EL's ELP.

### Writing Curriculum-Based Measurement (CBM)

CBM (Deno, 1985) has been recommended for screening and progress monitoring ELs to promote early identification of risk and access to early intervention (NASEM, 2017). CBM has been relatively well examined in reading for ELs, but the research examining Writing CBM with ELs is sparse (Keller-Margulis et al., 2016). A variety of forms of Writing CBM exist at the passage, sentence, and word levels of composition, but sentence-level Writing CBM, such as Picture Word (CBM-PW; McMaster et al., 2014), may have promise specifically for young ELs. CBM-PW assesses sentence and word-level transcription and text generation skills by presenting the student with 12 picture-word combinations while asking them to write their best sentence for as many picture-word combinations as they can in 3 min. CBM-PW has evidence of technical adequacy for students (general population) in the first through third grades (reliability:  $r \geq .70$ , validity:  $r \geq .50$ ; McMaster et al., 2011). Weekly prompts have also produced reliable slopes within 8 weeks and are sensitive to growth (McMaster et al., 2011).

CBM-PW may be effective specifically for young ELs because it is designed to capture writing growth in transcription and text generation at the word and sentence levels, which have been shown to be significant predictors of an EL's overall writing development (Babayigit, 2015; Harrison et al., 2016). In addition, CBM-PW provides scaffolds in the form of pictures and keywords that may promote access for ELs with beginning levels of ELP. CBM-PW is scored using a variety of metrics that may be roughly categorized as production (i.e., total words written [WW]), accurate-production (i.e., total words spelled correctly [WSC], total correct word sequences [CWS], and correct minus incorrect word sequences [C-IWS]),

production-independent (i.e., %WSC and %CWS; McMaster et al., 2014), and complexity (i.e., correct word sequences per response [CWSR]; Wagner et al., 2018). CWS is defined as two adjacent words that are both spelled correctly and used appropriately within the context of the sentence, inclusive of capitalization and punctuation (McMaster et al., 2014). CWSR is calculated by dividing the number of attempted sentences, with each sentence considered a response, by CWS to produce an average number of CWS per sentence/response (Wagner et al., 2018).

Prior research with the general population has indicated that accurate-production metrics produce the highest validity coefficients and best sensitivity to growth (McMaster et al., 2011). However, a prior study employing passage-level Writing CBM with advanced ELP ELs in fourth grade indicated promise for production-independent metrics (i.e., %CWS; Keller-Margulis et al., 2016). No prior study has examined CWSR with ELs. Reading, writing, and oral language are highly interrelated, but an important component of an assessment's validity is its ability to differentiate between highly related constructs (Salvia et al., 2017). CBM-PW's ability to discriminate between writing and oral language is especially important for ELs because they may be misidentified as being at risk or having a disability due to academic skills that are more directly related to ELP and the language acquisition process than to academic difficulties or disability (Abedi & Gándara, 2006; August & Shanahan, 2006). This study represents an initial examination of the technical adequacy of CBM-PW and its various metrics for ELs receiving English-only instruction in Grades 1, 2, and 3 with predominately intermediate to beginning ELP. The criterion measure employed is a common ELP assessment developed specifically for and normed with ELs that has subtests across the domains of reading, writing, and oral language, which allow for the examination of divergent validity (a component of validity). We sought to answer the following research questions:

**Research Question 1:** What is the reliability of CBM-PW metrics with young ELs?

**Research Question 2:** What is the predictive validity of CBM-PW metrics for young ELs?

## Method

### Setting and Participants

One midwestern U.S. school district serving a midsized city agreed to participate in the study. The district served 18,000 K–12 students: 60.8% were White, 20% Black, 6.3% Hispanic, and 5.5% Asian. Also, 39.7% of students were eligible for free and reduced-price lunch, 9.7% had Individualized Education Programs, and 6% received EL services. The estimated total number of EL students in

Grades 1, 2, and 3 across the district was 380, and parental consent letters were sent home to 230. Consent letters were provided in both English and the recipient's native language. All participants received English-only instruction. The total number of participants was 73 ELs across Grades 1 through 3. More sample information may be found in supplemental Table S1. ELP was assigned as beginning, intermediate, or advanced according to their ELP score with the majority in the beginning to intermediate range of ELP.

### Measures and Procedures

Two forms of CBM-PW created by McMaster and colleagues (2014) were used as predictor measures. Students took two forms (A, B) of CBM-PW in the fall (mid-November), winter (late-January), and spring (mid-April). The forms were counterbalanced across students and time points. CBM-PW was administered within a 2-week window across teachers and schools for each benchmark. The Assessing Comprehension and Communication in English State-to-State (ACCESS) English Language Proficiency test is the most common ELP assessment (administered in winter) across states (Fox & Fairbairn, 2011). The ACCESS-Writing subtest was used as the primary criterion measure in this study, and the Reading subtest and Oral Language Composite (OLC), a composite of the Receptive and Expressive oral language subtests, were used to examine divergent validity. Approximately 10% of all CBM-PW administrations per administrator were evaluated for fidelity of administration by a trained observer using a modified version of the Accuracy of Implementation Rating Scales (Fuchs et al., 1984). Total fidelity of administration was 99%. All scored data were double-counted and entered, and any discrepancies were discussed and remediated on an individual basis.

### Data Analysis

A random selection of 24% of all CBM-PW were double-scored for Inter-Scorer Reliability (ISR), counterbalanced across forms. ISR was calculated by dividing total scoring agreements by agreements plus disagreements. Pearson's correlations were used between forms of CBM-PW within each grade for each time point to examine alternate form reliability. According to Salvia et al. (2017), coefficients of  $r \geq .80$  are necessary for screening purposes. Predictive validity was examined using Pearson's correlations within each grade between mean CBM-PW metric scores and the ACCESS subtests of Writing, Reading, and the OLC.

## Results

### Descriptive Statistics

For the ACCESS-Writing, the Shapiro-Wilk test indicated a normal distribution ( $p > .05$ ) for Grades 2 and 3 but not for

Grade 1 ( $p < .05$ ). Follow-up analysis indicated an extreme outlier in first grade. The Shapiro-Wilk test indicated a normal distribution once the outlier was removed ( $p = .76$ ). CBM-PW descriptive statistics are provided in the supplemental Table S2 with the outlier removed. All subsequent analyses were conducted with the outlier removed.

### Reliability and Validity

The ISR across metrics ranged from 88% to 98%, with C-IWS consistently having the lowest reliability while the others were consistently above 90%. Alternate form reliability for CBM-PW is available in Table 1. Predictive validity results are in Table 2. The full correlation tables for all metrics, including predictive and concurrent validity, are available in supplemental Table S3.

## Discussion

The purpose of this study was to examine the reliability and validity of various CBM-PW metrics for young ELs receiving English-only instruction with predominately intermediate to beginning ELP. The findings indicated that word-level skills (i.e., spelling and word generation) as measured by WW and WSC were the most consistent predictors of writing performance as measured by the ACCESS-Writing. WW and WSC demonstrated the highest levels of adequacy for second-grade ELs, but no metric met appropriate levels of adequacy for first grade. For first grade, students may not yet have the English literacy and language skills to be assessed adequately with CBM-PW. The primary metrics (i.e., WW, WSC, and CWS) generally correlated more strongly with ACCESS-Writing than the ACCESS-OLC for Grades 1 and 2, but correlations were equitable across subtests for Grade 3. Clearly, more research is needed to examine the relationship(s) between oral English proficiency and the various CBM-PW metrics.

Although the findings here are generally in line with prior research examining CBM-PW with the general population (i.e., technical adequacy is stronger for second grade than for first grade; McMaster et al., 2011), there were a couple of interesting contradictions between this study and prior studies (i.e., Keller-Margulis et al., 2016; McMaster et al., 2011). In contrast to CBM-PW studies with the general population (McMaster et al., 2011; Wagner et al., 2018), metrics incorporating sentence-level features in this study (i.e., CWS, C-IWS, CWSR) did not perform as well or as consistently as word-level metrics (i.e., WW, WSC). This could be a distinct feature of EL writing progression, an anomaly distinct to this study's sample, or a feature specific to the criterion measure (ACCESS). Also, production-independent metrics (i.e., %WSC, %CWS) that showed promise in a prior study with older ELs employing passage-level Writing CBMs (Keller-Margulis et al., 2016) produced no significant

**Table 1.** Alternate Form Reliability for Picture Word.

Metric	Fall			Winter			Spring		
	1st (n = 21)	2nd (n = 23)	3rd (n = 23)	1st (n = 23)	2nd (n = 25)	3rd (n = 23)	1st (n = 20)	2nd (n = 25)	3rd (n = 22)
WW	.81*	.90*	.91*	.88*	.94*	.81*	.77*	.91*	.97*
WSC	.72*	.89*	.92*	.85*	.92*	.83*	.71*	.88*	.96*
CWS	.60*	.72*	.88*	.91*	.82*	.84*	.86*	.88*	.94*
IWS	.75*	.86*	.61*	.54*	.70*	.48*	.91*	.78*	.71*
%WSC	.62*	.35	.49*	.68*	.73*	.67*	.70*	.31	.69*
%CWS	.42	.65*	.58*	.86*	.70*	.73*	.85*	.62*	.66*
C-IWS	.30	.45*	.77*	.84*	.65*	.84*	.91*	.81*	.88*
CWSR	.46*	.84*	.84*	.77*	.94*	.77*	.80*	.89*	.30

Note. 1st = first grade; 2nd = second grade; 3rd = third grade; WW = words written; WSC = words spelled correctly; CWS = correct word sequences; IWS = incorrect word sequences; C-IWS = correct minus incorrect word sequences; CWSR = correct word sequences per response. \*p < .05.

**Table 2.** Predictive Validity of PW in Fall With ACCESS Subtests of Writ, Read, and OLC in Winter for Grades 1, 2, and 3.

Grade (n)	First grade (n = 20)			Second grade (n = 22)			Third grade (n = 23)		
	Writ	Read	OLC	Writ	Read	OLC	Writ	Read	OLC
PW metrics									
WW	.43	.38	.30	.61*	.46*	.02	.53*	.43*	.52*
WSC	.42	.49*	.16	.62*	.49*	.01	.56*	.44*	.48*
CWS	.38	.45*	.15	.65*	.61*	.20	.56*	.45*	.50*
%WSC	-.01	.21	-.25	.14	.40	-.11	.32	.09	-.32
%CWS	.06	.03	-.10	.24	.33	.56*	.34	.26	.25
C-IWS	.09	.18	-.10	.44*	.57*	.45*	.49*	.41	.40
CWSR	-.1	-.23	-.09	.56*	.34	.46*	.37	.35	.45*

Note. PW = Picture Word; ACCESS = Assessing Comprehension and Communication in English State-to-State English Language Proficiency; OLC = Oral Language Composite; Writ = Writing; Read = Reading, WW = words written; WSC = words spelled correctly; CWS = correct word sequences; C-IWS = correct minus incorrect word sequences; CWSR = correct word sequences per response. \*p < .05.

correlations with the ACCESS-Writing in this study. This could mean that accuracy, which production-independent metrics (e.g., %WSC) predominately assess, is not as predictive for young ELs or ELs with intermediate to beginning ELP as it is for older ELs (fourth grade) with more advanced ELP. These findings provide additional evidence that assessment tools should be examined and validated specifically for ELs according to ELP and grade level.

**Limitations and Future Research**

Results should be considered exploratory and may not generalize to ELs beyond the sample included in this study. The sample size for this study did not allow for analyses based on ELP; therefore, it is unclear whether CBM-PW is able to identify risk for ELs within each level of ELP. For example, ELs with beginning ELP should perform and grow at different rates than ELs with advanced ELP due to typical language acquisition processes. If risk is defined as

unexpected or atypical underperformance, then it is essential to establish what expected or typical performance is for each level of ELP. Therefore, future studies should recruit sample sizes large enough to examine CBM-PW across both grades and ELP (including advanced ELP). Furthermore, participants in this study received English-only instruction and, in accordance with research indicating that ELs progress differently when receiving bilingual instruction (August & Shanahan, 2006), future studies should examine CBM-PW performance specifically for ELs receiving bilingual instruction.

Beyond limitations with sample size and characteristics, this study only employed one criterion measure. Although the ACCESS is used to determine ELP and the nature of English as Second or Other Language services a student receives, a variety of criterion measures should be used to examine assessment validity (Salvia et al., 2017). Future research may also examine the impact native language proficiency has on CBM-PW. Finally, studies should examine



new avenues of scoring CBM-PW for young ELs that account for vocabulary and syntactic complexity because these two constructs are critical to ELs' literacy development (August & Shanahan, 2006; Babayiğit, 2015; Harrison et al., 2016) but may not be adequately captured by existing CBM-PW metrics.

### Implications

The real value of CBM-PW may lie more in the domain of instructional utility than in technical adequacy as a screener (especially given the difficulties inherent in assessing writing; Salvia et al., 2017). For example, teachers may be able to examine performance across word-level (e.g., WSC), sentence-level (e.g., CWS), and complexity (e.g., CWSR) metrics to identify areas of need for instructional focus (spelling/vocabulary, grammar, or sentence combining, respectively). Furthermore, using CBM-PW on a regular (weekly or biweekly) basis may encourage teachers to increase their writing instruction in general, and the sentences produced by students provide rich sources of data from which teachers can identify student strengths and needs. Finally, repeated practice in generating sentences as is necessary with CBM-PW may promote sentence writing fluency.

### Authors' Note

The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

### Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A130144 to the University of Minnesota.

### Supplemental Material

Supplemental material for this article is available on the *Assessment for Effective Interventions* website with the online version of this article.

### References

- Abedi, J., & Gándara, P. (2006). Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practice*, 25(4), 36–46.
- August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second language learners: A report of the National Literacy Panel on Language Minority Children and Youth*, Center for Applied Linguistics. Lawrence Erlbaum.
- Babayiğit, S. (2015). The dimensions of written expression: Language group and gender differences. *Learning and Instruction*, 35, 33–41.
- Berninger, V. W., & Amtmann, D. (2003). Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: Research into practice. In H. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 323–344). Guilford Press.
- Deno, S. L. (1985). Curriculum-Based Measurement: The emerging alternative. *Exceptional Children*, 52, 219–232.
- Fox, J., & Fairbairn, S. (2011). Test review: ACCESS for ELLs [R]. *Language Testing*, 28(3), 425–431.
- Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent Curriculum-Based Measurement and evaluation on pedagogy, student achievement, and student awareness of learn-in. *American Educational Research Journal*, 21, 449–460.
- Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, 81(4), 710–744.
- Harrison, G. L., Goegan, L. D., Jalbert, R., McManus, K., Sinclair, K., & Spurling, J. (2016). Predictors of spelling and writing skills in first and second-language learners. *Reading and Writing*, 29(1), 69–89.
- Keller-Margulis, M., Payan, A., Jaspers, K. E., & Brewton, C. (2016). Validity and diagnostic accuracy of written expression Curriculum-Based Measurement for students with diverse language backgrounds. *Reading & Writing Quarterly*, 32(2), 174–198.
- McMaster, K. L., Du, X., Yeo, S., Deno, S. L., Parker, D., & Ellis, T. (2011). Curriculum-based measures of beginning writing: Technical features of the slope. *Exceptional Children*, 77(2), 185–206.
- McMaster, K. L., Lembke, E., Brandes, D., Garman, C., Moore, K., Jung, P., & Janda, B. (2014). *Data-based instruction in beginning writing: A manual*. Department of Educational Psychology, University of Minnesota.
- National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the educational success of children and youth learning English: Promising futures*. The National Academies Press. <https://doi.org/10.17226/24677>
- Roseberry-McKibbin, C., & Brice, A. (n.d.). *Acquiring English as a second language: What's "normal," what's not*. <http://www.asha.org/public/speech/development/easl/>
- Salvia, J., Ysseldyke, J., & Witmer, S. (2017). *Assessment in special and inclusive education* (13th ed.). Cengage Learning.
- Wagner, K., Smith, R. A., Allen, A., McMaster, K., Poch, A., & Lembke, E. (2018). Exploration of new complexity metrics for Curriculum-Based Measures of Writing. *Assessment for Effective Intervention*, 44(4), 256–266. <https://doi.org/10.1177/1534508418773448>