Psicología Educativa

Curriculum-based Measurement for Early Writing Struggles in Kindergarten: A Systematic Review

Jennifer Balade[1], Cristina Rodríguez[2,3], and Juan E. Jiménez[1]

[1]Universidad de La Laguna, The Canary Islands, Spain, [2]Universidad de Talca, Chile
[3]Millennium Nucleus for the Science of Learning, Talca, Chile

ABSTRACT

Early detection and intervention play a crucial role in the prognosis of children with specific learning disabilities. Due to the complexity of the writing process and its persistence into adulthood, tools for early detection are essential. Curriculum-based measures (CBMs) are quick, reliable, and evidence-based tools used for early assessments. This systematic review aims to summarize the technical features of CBM in writing for kindergarten students following the PRISMA guidelines. A total of 311 articles from PsycINFO, SCOPUS, ERIC, and WOS were examined, and five studies met the inclusion criteria. Production measures and transcription skills were the most studied. The reviewed tasks demonstrated criterion validity, reliability, sensitivity to the timing of assessment, or diagnostic accuracy in at least one of their measures, with a greater presence of the dependent production measures. The findings emphasize the scarcity of evidence-based tools for identifying early writing difficulty, underscoring the need for rigorous research in early childhood education.

Medidas basadas en el currículo para dificultades tempranas de escritura en Educación Infantil: una revisión sistemática

RESUMEN

La detección e intervención tempranas juegan un papel crucial en el pronóstico de los niños con dificultades específicas de aprendizaje. Debido a la complejidad de los procesos de escritura y su persistencia en la edad adulta, las herramientas para la detección precoz en escritura son esenciales. Las medidas basadas en el currículo (MBC) son herramientas rápidas, fiables y basadas en la evidencia utilizadas para evaluaciones tempranas. Esta revisión sistemática tiene como objetivo resumir las características técnicas de las MBC en la escritura para estudiantes de Educación Infantil siguiendo las directrices PRISMA. Se examinaron un total de 311 artículos de PsycINFO, SCOPUS, ERIC y WOS y cinco estudios cumplían con los criterios de inclusión. Las MBC de producción y las habilidades de transcripción fueron las más estudiadas. Las tareas revisadas demostraron validez de criterio, fiabilidad, sensibilidad en el momento de la evaluación o precisión diagnóstica en al menos una de sus medidas, con una mayor presencia de las medidas de producción dependientes. Los resultados destacan la escasez de herramientas basadas en pruebas para identificar dificultades tempranas en la escritura, subrayando la necesidad de una investigación rigurosa en educación infantil.

Writing is a feature of daily life (Graham & Hall, 2016), so the development of this ability is crucial. However, writing is one of the most complex academic skills to learn, given the degree of cognitive processing involved (Graham et al., 2013). Extensive research has shown that writing requires lower-order transcription skills, such as spelling, alphabetic knowledge, grapheme-phoneme correspondence, and graphomotor skills (Berninger & Amtmann, 2003; Graham et al., 1997; Graham & Harris, 2005; Juel et al., 1986); higher-order

processes of text composition, such as planning, idea generation, and establishing objectives (Berninger & Amtmann, 2003; Flower & Hayes, 1981; Graham et al., 1997); memory skills, such as long-term memory and working memory (Berninger & Winn, 2006); and oral language skills (Berninger, 2000; Berninger & Winn, 2006; Hooper et al., 2011; Kent et al., 2014; Kim et al., 2011; Rodríguez et al., 2024).

Because of the complexity of writing, some students experience difficulty mastering it. Children with poor writing proficiency display

diminished self-efficacy and reduced motivation to acquire or enhance their writing abilities (Graham et al., 2017). Additionally, writing disabilities may persist into adulthood (MacArthur & Philippakos, 2013). As such, low performance not only causes academic problems but it can also affect individuals' future workplace experiences or even their emotional well-being (Graham & Hall, 2016).

According to the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V; American Psychiatric Association, 2013), the prevalence of specific learning disabilities (SLDs), including reading, mathematics and writing impairments, is approximately 5% to 15%. Focusing on writing, several studies have reported the prevalence of writing disabilities. Research conducted in languages used by Western cultures (including Europe, the United States, and Latin America) has shown that the prevalence of writing disabilities ranges from 3% to 16.4% (Bosch et al., 2021; Cappa et al., 2015; Fortes et al., 2016; Jiménez & García de la Cadena, 2007; Katusic et al., 2009; Landerl & Moll, 2010; Leonardi et al., 2021; Moll et al., 2014; Vélez Calvo, 2017; Von Aster et al., 2010). These figures underscore the worldwide prevalence of the problem and its effect on speakers of different languages, regardless of orthographic transparency.

Existing research recognizes the critical role played by early identification in SLDs. As soon as the problem is identified, intervention is implemented, providing a more favorable prognosis (Fuchs & Fuchs, 2006; Fuchs & Fuchs, 2007; Parker et al., 2012). However, SLD tends to be underidentified in the early grades and overidentified in the upper grades of elementary school (National Academies of Sciences, Engineering, and Medicine, 2017). In addition, early detection of SLDs has positive effects on self-esteem and motivation to write among these students (Blanco Pérez & Bermejo, 2009; Defior et al., 2015). Focusing on writing, previous studies have shown that early identification and intervention in individuals with these disabilities prevent long-term negative consequences in terms of writing performance, thus mitigating negative consequences in daily life as adults (Berninger & Amtmann, 2003; Berninger et al., 2008; Hooper et al., 2011). Therefore, it is especially important to develop instruments that allow for the assessment of writing from the early years of schooling. In this regard, we have writing assessment instruments in Spanish, but they have primarily focused on students starting from the first grade of primary education (Jiménez, 2018). While the Early Grade Writing Assessment (EGWA) is a useful tool, it was designed for primary education and not for kindergarten. In the present study, however, the focus is on kindergarten.

Despite the prevalence of writing disabilities and their long-term consequences, there is a relative scarcity of research dedicated explicitly to the investigation of writing disabilities, in contrast to the wealth of studies focusing on reading or mathematics disabilities (Arrimada et al., 2020; Berninger et al., 2008). Some authors have even characterized writing disabilities as "the forgotten learning disability" (Katusic et al., 2009), underscoring how the issue is often neglected in research. This lack of focus has significant implications, notably contributing to delayed identification (Graham & Harris, 2005), which in turn can limit educational and employment opportunities (Graham & Perin, 2007).

Response to Intervention and Curriculum-based Measurement

The response to intervention model (RtI) is a preventive model for the early detection of SLD that is widely accepted in the scientific literature. The model has four main elements: 1) a multilevel prevention system, 2) universal screening to detect students at risk of learning disabilities, 3) monitoring the learning progress of at-risk students, and 4) decision making based on objective and standardized data (National Center on Response to Intervention, 2010). In the

RtI context, curriculum-based measures (CBM) are often used for universal screening and progress monitoring. These measures are assessed at different points throughout the year (normally in autumn, winter, and spring) to detect at-risk students and decide whether the intervention should be modified or if progress is acceptable (Johnson et al., 2006). CBMs consider the basic competencies established in the school curriculum (Hosp et al., 2007). These tools are quick and practical for identifying at-risk students and have high levels of reliability and validity. They are also sensitive to learning growth over time, allowing evaluation of the effectiveness of the intervention (Ardoin, 2006; de León et al 2021a; de León et al., 2021b; Gutiérrez et al., 2021; Jiménez et al., 2021).

Many writing tasks and scores in early grades are focused on transcription skills (McMaster et al., 2011). Extensive research has shown that transcription skills are the component most studied in beginning writers (Kirby et al., 2021), while higher-order processes have been more extensively explored in higher grades, as young students do not produce large volumes of text, and studies with samples of beginning writers present limitations when discourse measurements are taken (Ritchey et al., 2015). Several authors have shown that early transcription skills are the best predictors of writing success (Berninger et al., 2002; Coker, 2006; Graham et al., 1997; Kim et al., 2011; Tortorelli et al., 2022).

CBM in writing (CBM-W) encompasses scores at multiple levels of language: subword, word, sentence, and discourse (Whitaker et al., 1994). First, students develop alphabetic knowledge and phoneme-grapheme correspondence to transcribe letters, sounds, and words (at the subword and word levels) (Ehri, 1986). Once writing conventions have been acquired, they begin to generate multiple words (sentence level) (Tolchinsky, 2006) and eventually produce longer units of text and ideas (discourse level) (McCutchen, 2006). Given the complexity of writing and the multiple levels on which to measure it, CBM scores have used two dimensions to capture writing performance: production-dependent scoring and production-independent scoring (Ritchey et al., 2015).

Focusing on these dimensions, the following scoring methods are used for CBM-W: 1) "production-dependent scoring", which measures the quantity of text written with indices such as words written (WW), words spelled correctly (WSP), correct word sequence (CWS), correct minus incorrect word sequences (CIWS) (Deno et al., 1982; Parker et al., 2009; Videen et al., 1982), correct letter sequences (CLS) (McMaster & Espin, 2007), number of long words (Coker & Ritchey, 2010; Lembke et al., 2016; McMaster et al., 2011; McMaster & Espin, 2007), or number of complete sentences (Gansle et al., 2002, 2006). The reliability, validity, and sensitivity at the time of assessment vary depending on the age of the students and the type of task used (Arrimada et al., 2022; McMaster & Espin, 2007); 2) "production-independent scoring", prioritizing quality over quantity; other authors also regard these measures as precision measures (Arrimada et al., 2022); this score includes the percentage of words spelled correctly (%WSC) and the percentage of correct word sequences (%CWS) (Jewell & Malecki, 2005; Parker et al., 2009). Similar to production-dependent measures, this type of measurement exhibits varying reliability, validity, and sensitivity at the time of assessment depending on the school age at which it is evaluated, as well as the type of prompt used (Arrimada et al., 2022).

Measurement of the above scores requires quick, useful, and easy-to-score tasks that serve as indicators of writing proficiency. As reported by Ritchey et al. (2015), the following CBM-W tasks have been proposed in early grades (prekindergarten and kindergarten): 1) "name writing", where students had to write their own names. This task appears to have moderate validity ratings ($r = .36$) in kindergarten (Lonigan et al., 2008), but some authors consider it inadequate, as children write their name from memory rather than through transcription processes. In addition, validity indices lose robustness as students reach primary education due to ceiling

effects (Puranik & Lonigan, 2012); 2) "letter copying", where a sheet of paper with capital letters is handed out to students and they are instructed to copy each letter in 1 minute. However, in this task, we find moderate reliability indices (r = .68), but validity indices range from low to moderate (r = .21-.59), depending on the test criteria. The same authors concluded that the task could be used in the early years of early childhood education (VanDerHeyden et al., 2001); 3) "letter writing fluency", where students were asked to write lowercase letters of the alphabet in 1 minute. This task exhibited moderate validity indices (r = .44-.58). The authors propose this task as a potential option for use with kindergarten students (Ritchey et al., 2015); 4) "word dictation", where words are read twice once students have finished the previous word or after a pause. This task showed reliability indices ranging from moderate to high (r = .55-.97) depending on the task duration and moderate validity (r = .48-.50), making it a suitable word-level task for kindergarten (Hampton et al., 2012); 5) "sentence writing", where an oral prompt about a topic was presented and students had to generate text in 3 minutes. The reliability indices of this task were high (r = .74 - .87), although the validity indices were low (r = .20-.46) for the kindergarten sample (Coker & Ritchey, 2010).

Although CBM are crucial for the early identification of children at risk of writing difficulties, there is limited prior literature verifying the reliability, validity, and growth sensitivity of CBM in writing. Furthermore, the majority of CBM-W have been studied at primary and secondary school ages (Arrimada et al., 2020; McMaster & Espin, 2007). The National Center on Intensive Intervention website does not feature academic screening tools for writing, even though tools for reading and mathematics are available (The National Center on Intensive Intervention, 2021).

To our knowledge, only two systematic reviews have been conducted. The review conducted by McMaster and Espin (2007) provides a comprehensive summary of the technical features of available CBM-W, including validity and reliability indices of the reviewed measures. These authors concluded that simple production-based measures, such as total words written or total sentences written, are more effective at early ages, while more complex measures, such as the percentage of correct words or complexity of written ideas, are more suitable and technically stronger in secondary education. Nevertheless, the review was conducted for the elementary, middle, and high school levels but did not encompass kindergarten. Another more recent review conducted by Arrimada et al. (2020) examined writing assessment measures, including CBM. The authors identified a total of eleven articles analyzing CBM-W based on production scores and eight articles analyzing CBM-W based on precision scores. Although this review does not examine psychometric characteristics such as the reliability, validity, diagnostic accuracy, or growth sensitivity of the measures, it provides information on the types of CBM utilized, prompt types, and benefits and limitations of the measures. Another conclusion drawn is that production-based measures are more effective in the early years of elementary school, while precision measures are more effective in later years of elementary and high school, in line with the findings of McMaster and Espin (2007). However, all the articles reviewed in this study involved participants from primary and secondary education levels, without including studies conducted in kindergarten. Consequently, an updated and focused perspective on CBM-W focused on early childhood education is currently lacking.

The purpose of the present systematic review is to compile the CBM-W assessed in the kindergarten population and to summarize its psychometric characteristics, such as reliability, validity, diagnostic accuracy, and sensitivity to the growth of these measures. In addition, this review aims to provide an overview of the utility and efficacy of these measures in the kindergarten context.

Method

This systematic review was performed according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, version 2020 (Page et al., 2021). The PRISMA guidelines were followed to ensure a comprehensive and transparent approach to the review process.

Search Strategy

We searched for relevant documents from four electronic databases: SCOPUS, WOS, PsycINFO, and ERIC. To ensure comprehensive coverage, we employed various search techniques. The asterisk truncation wildcard (*) was used to find all documents containing a specific word and its possible endings. We also used quotation marks (" ") to find specific two-word concepts (e.g., one concept commonly used by researchers is "classification accuracy") as opposed to the two words separately.

The search formula contains four topics. 1) The first one involves words relating to *psychometric characteristics*. This topic was as follows: (screen* OR benchmark* OR "diagnostic accuracy" OR "classification accuracy" OR utilit* OR "technical adequacy" OR roc). 2) The second topic encompassed words relating to tasks or processes involved in "written expression". We used ("written expression" OR "story prompt" OR "T-Unit" OR "written letters" OR "word copy" OR "sentence copy" OR "word dictation" OR spelling OR handwriting). 3) The third topic specifies that the writing measures must be "curriculum-based measures". We added (CBM OR "curriculum-based measure*"); 4) The last topic contains words concerning the "participants' age". We were interested in finding documents with the youngest possible samples. We used (kinder* OR "nursery schools" OR "day nurseries" OR "early childhood education" OR "preschool children"). The last four concepts were taken from the Thesaurus as proposed synonyms of the word "kindergarten" United Nations Educational, Scientific, and Cultural Organization [UNESCO, 1977].

The four topics (psychometric characteristics, writing expression, CBM, and kindergarten samples) were united by the logical operator AND in the search formula. The final formula incorporating the four topics was applied across the four aforementioned databases through the "advanced search" section. The systematic search yielded 322 documents (including duplicates). Seven further documents were added by one of the authors based on her prior awareness of them and her compliance with the search criteria. Thus, a total of 329 documents were included in the next step.

Inclusion and Exclusion Criteria

Once 18 duplicate articles had been removed, document titles and abstracts were screened for inclusion in the present study. Publications were only included in the screening if they 1) reported empirical articles, 2) involved the evaluation of CBM-W psychometric characteristics, 3) applied in kindergarten, and 4) were published in English or Spanish (from any country). The exclusion criteria were as follows: 1) the paper was not reported as an empirical study, 2) the paper was not intended to evaluate the psychometric characteristics of the CBM-W (e.g., whether they used these measures in an intervention context), 3) the study sample was students in grades above kindergarten, and 4) the studies were published in languages other than English and Spanish. A total of 311 articles were subjected to the screening procedure.

Screening Process

The screening process was conducted according to the method described by Belur et al. (2021), which entails screening documents

Table 1. Comparative IRR Test Scores and κ Statistic

Categories	First IRR		Second IRR		Third IRR	
	No Studies	Percent	No Studies	Percent	No Studies	Percent
Total items coded individually	103	100	104	100	103	100
Screening decision—include[1]	45		23		27	
Screening decision—exclude	264		290		282	
Individual coding—agreement for all[2]	98	95.14	103	99.03	103	100
Joint coding—agreement for all[3]	5	4.85	1	0.96	0	-
Lack of agreement after discussion	0		0		0	
Agreement after arbitration	NA	NA	NA	NA	NA	NA
κ statistic[4]	.87		.95		1	

Note. Table design based on the study by Belur et al. (2021); IRR = interrater reliability.
[1]Total number of screening decisions made by the three judges, both included and excluded.
[2]Total number of agreements included and excluded by the three judges.
[3]Total number of studies that required discussion regarding inclusion or exclusion.
[4]κ statistic was calculated based upon the number of inclusions and exclusions after the initial screening decision and prior to reconciliation.
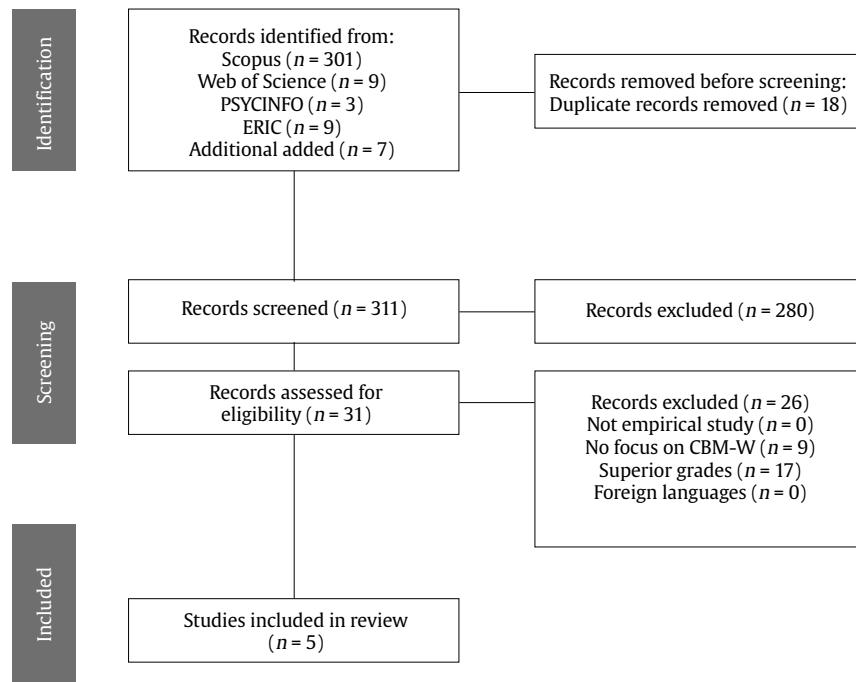

Figure 1. PRISMA Flow Diagram.

in 3 blocks. In each block, the judges conduct their own screening and then discuss agreements and disagreements before calculating the IRR. This method allows judges to refine their discussion in each block and enhance their understanding of the inclusion criteria, thereby improving the IRR indices. Likewise, we used the Rayyan tool to classify articles as included or excluded for the screening procedure. This tool facilitates the independent screening of documents by multiple judges, ensuring that they are unaware of the screening procedures made by their peers (Ouzzani et al., 2016).

Following the described methodology, we divided the 311 articles into 3 batches or blocks. The first block contained the initial 103 articles (in alphabetical order) that underwent blind screening by the three judges. After this process, the interrater reliability (IRR) among the three judges was calculated, and disagreements regarding article inclusion-exclusion based on established criteria were discussed. Following this discussion, the inclusion and exclusion criteria were further detailed to prevent subsequent

discrepancies in the following screening blocks. Subsequently, the screening process for the second block was conducted, which included 104 articles, representing the subsequent entries in alphabetical order for review. Blind screening was performed by the three judges, the IRRs were calculated, and disagreements regarding article inclusion or exclusion were discussed upon completion. Once discussed and justified, the inclusion and exclusion criteria were further delineated, and the screening process for the final block was initiated. The third block of articles consisted of the last 103 articles. The same procedure as before was followed, where blind screening was conducted, the IRR was calculated, and in this block there was no need to discuss disagreements, as there were no disagreements (see Table 1). IRR indices were computed using Fleiss' kappa in each block, achieving .87 in the first block, .95 in the second block, and a perfect score of 1.0 in the third block (see Table 1). Kappa values approach perfection according to the interpretations by Landis and Koch (1977).

Data Collection Process

The blockwise screening process yielded 31 articles. To conduct the data collection process, the full texts of these documents were obtained. The first author examined the complete text of each document, while the other two authors reviewed them to verify that they met the criteria. Any disagreements were resolved through discussion. Twenty-six documents were excluded during this eligibility phase. The inclusion and exclusion criteria for this process mirrored those of the screening process. Seventeen articles were excluded for utilizing student samples beyond the kindergarten level, and an additional 9 articles were excluded for failing to assess the psychometric properties of the CBM-W. In the end, only 5 articles presented the characteristics we were looking for. These were all then included in the review (see Figure 1).

Data Item Process

A total of 5 articles were manually coded by the first author. The documents were coded according to the following characteristics: 1) study information (i.e., authors and year), 2) sample information (e.g., country of study, native language, and school grade), 3) writing measures (including information about CBMs, type of prompt, task duration, moment of measurement, type of measure, and scoring procedures), 4) criterion validity (including criterion measure and correlation coefficient), 5) reliability (results in alternate form, internal consistency, and interrater reliability), 6) sensitivity to measurement time (slopes of the fixed- effect estimates), and 7) diagnostic accuracy (including sensitivity, specificity, and area under the curve [AUC]). See Table 2.

Study Quality

The methodological quality and biases of the included studies were assessed using the McMaster Critical Review Form - Quantitative Studies (CRF-QS). In each study, 15 items were evaluated, which were divided into 8 main domains: 1) study purpose, 2) literature, 3) study design, 4) sample, 5) outcomes, 6) intervention, 7) results, and 8) conclusions and implications. A score of 1 was assigned when a criterion was met, a score of 0 was given if the criterion was not met, and a score of N/A was assigned if the criterion was not applicable (Law et al., 1998). We adapted the tool by excluding the intervention section since it was not applicable to the studies in our review (see Table 3). Therefore, the maximum score for our studies was 12. Using the scoring guidelines provided by Faber et al. (2015), a final score expressed as a percentage was calculated for each study. Finally, based on the percentage of the final scores, each study was classified as follows: 1) low methodological quality, with a score ≤ 50%, 2) good methodological quality, with a score between 51 and 75%, and 3) excellent methodological quality, with a score > 75%.

Table 2. Summary of Study Characteristics

Study	Sample	Writing measures — Type of prompt	Task duration	Moment(s) of measurement	CBM	Criterion validity — Criterion measure(s)	Correlation coefficient(s)	Reliability — Alternate-form	IRR	Sensitivity to growth	Diagnostic accuracy — Sensitivity	Specificity	AUC
Coker & Ritchey (2010)	76; United States; English; Kinder and 1st	Sentence writing	3 min	January, March, and May	TWW CSP CWS	TEWL-2: Basic writing and Contextual writing subtests	.36**-.46*** .24*-.30** .20-.23**	.76 .79 .78	.99 .98 .97	1.50***1 1.38***1 1.10***1			
Justi et al. (2020)	213; Brazil; Portuguese; Kinder	LNK-15 items LNK-26 items	2 min		CL-15 CL-26	WPT1					.80 .80	.82 .82	.60 .50
Keller-Margulis et al. (2019)	96; United States; English; Kinder	Word dictation Picture word			TWW CSW TLW CLS TWW CSW CWS TLW CLS	WJ-IV: Writing expression, writing samples and sentence writing fluency subtests	.32**-.30*-.26** .70*-.60*-.61* .47*-.42*-.41* .68*-.59*-.56* .48*-.40*-.51* .53*-.42*.57* .56*-.45*.60* .49*-.43*-.48* .61*-.50*-.61*		.98-1 .86-1		.75 1 .75 .94 .69 .69 .81 .75 .69	.58 .82 .71 .81 .64 .60 .77 .53 .76	.75 .92 .78 .93 .74 .76 .83 .69 .83
Puranik et al. (2017)	134; United States; English; Kinder	Alphabetic writing fluency	untimed	At the beginning of the course At the end of the course	CL 15 s CL 60 s CL-UNT CL 15s CL 60s CL-UNT	TEWL-3, WJ-III, Picture word and Essay prompt	.21-.24-.41**-.31** .46**-.48**-.39**-.39** .47**-.55**-.63**-.58** .46**-.51**-.58**-.58** .49**-.55**-.64**-.59** .45**-.57**-.54**-.59**						
Ritchey et al. (2010) Study 2	76; United States; English; Kinder	Word Dictation		January, March, and May	CSP CLS CS PC	TEWL-2: Basic writing and Contextual writing subtests	.54**-.26* .57**-.33** .59**-.36** .58**-.35**		99.6% 98.7% 95.1% 97.0%	.95*** 2.94*** 2.11*** 3.36***			

Note. AUC = Area under curve; CL = Correct letters; CL 15 = Correct letters 15 items version in Letter Name Knowledge task; CL 26 = Correct letters 26 items version in Letter Name Knowledge task; CL-15s = Correct letters in the first 15 seconds in Alphabetic Writing Fluency task; CL-60s = Correct letters in the first 60 seconds in Alphabetic Writing Fluency task; CL-UNT = Correct letters in total time (untimed) in Alphabetic Writing Fluency task; CLS = Correct letter sequences; CL = Correct letters; CSW = Correct spelled words; CWS = Correct word sequences; IRR = Interrater reliability; LNK-15/26 = Letter-name knowledge 15 items or 26 items task; PC = Phonological coding; TLW = Total letters written; TEWL-2/3 = Test of Early Written Language 2nd or 3rd edition; TWW = Total words written; WJ-III/IV = Woodcock Johnson Test of Achievement 3rd or 4th edition; WPT1 = Writing precision test 1st-grade.

[1] Analysis conducted with a combined sample of kindergarten and 1st grade students.

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

Table 3. Results of Methodological Quality Using the Critical Review Form – Quantitative Studies (Law et al., 1998)

Study	Study Purpose	Literature	Study Design	Sample		Outcomes			Results			Conclusions and Implications	Scores (%)	Interpretation
	1	2	3	4	5	6	7	8	9	10	11	12		
Coker and Ritchey (2010)	+	+	+	+	+	+	+	+	-	+	+	+	91.66	Excellent
Justi et al. (2020)	+	+	+	+	-	?	N/A	+	+	+	-	+	66.66	Good
Keller-Margulis et al. (2019)	+	+	+	+	-	+	+	+	+	+	-	+	83,3	Excellent
Puranik et al., (2017)	+	+	+	+	-	N/A	+	+	+	+	-	+	75	Good
Ritchey et al. (2010)	+	+	+	+	-	+	+	+	+	+	-	+	83,3	Excellent

Note. 1) clear study purpose; 2) relevant background literature reviewed; 3) design appropriate for the study question; 4) sample described in detail; 5) sample size justified and/or informed consent obtained; 6) reliable measures; 7) valid measures; 8) results reported in terms of statistical significance; 9) analysis method(s) appropriate; 10) clinical or educational importance reported; 11) drop-outs reported; 12) conclusions were appropriate given study methods and results; + = criterion is fully met; - = non-fulfilment; ? = unclear; N/A = not applicable.

Results

All the documents included in the systematic review were published in 2010 or later. As Table 2 shows, most of the studies were carried out with English-speaking students (*n* = 4), and one of them was conducted with Portuguese speakers (*n* = 1). The CBM measures in writing reported in this review were as follows: 1) Total Words Written (TWW), 2) Correctly Spelled Words (CSW), 3) Correct Word Sequences (CWS), 4) Correct Letters (CL), 5) Total Letters Written (TLW), 6) Correct Letter Sequences (CLS), 7) Phonological Coding (PC). These measures were scored within different tasks for the kindergarten population.

The tasks used to assess these measures were: 1) Sentence Writing, 2) Letter Name Knowledge, 3) Word Dictation, 4) Picture Word, and 5) Alphabetic Writing Fluency. Table 4 shows the frequency of CBM-W utilization across various tasks within different studies. Only the Word Dictation task was employed in two studies to assess CBM-W. The most studied CBM writing in kindergarten is Correctly Spelled Words, which appeared in 4 instances in the reviewed studies, followed by Total Words Written, Correct Letters, and Correct Letter Sequence measures, which were assessed 3 times. Correct Word Sequences and Total Letters Written measures were evaluated twice, while the Phonological Coding measure appeared only once. The validity, reliability, growth sensitivity, and diagnostic accuracy of each measure will then be examined.

Validity, Reliability, Growth Sensitivity, and Diagnostic Accuracy of the CBM-W

The Correctly Spelled Words (CSW)

This measure is evaluated in word dictation tasks and sentence generation tasks, such as picture word or sentence writing (Coker & Ritchey, 2010; Keller-Margulis et al., 2019; Ritchey et al., 2010). In relation to "validity", this measure shows significant correlation

coefficients ranging from .24 to .70 among the different studies for the TWEL-2 and WJ-IV criterion measures. Regarding "reliability" indices, the most frequently reported statistic is the IRR, with indices ranging from .98 to 1 (Coker & Ritchey, 2010; Keller-Margulis et al., 2019) or with 99.6% agreement between raters (Coker & Ritchey, 2010). Only one study reported the reliability of its alternate form, with a score of .77 (Coker & Ritchey, 2010). In terms of "growth sensitivity", this measure proved sensitive at different measurement points (January, March, and May), with estimated slope values of 0.95 (*p* = .0001) and 1.38 (*p* = .0001), although the latter was obtained through an analysis that included kindergarten and first-grade samples (Coker & Ritchey, 2010). Regarding "diagnostic accuracy", CSW demonstrated effective diagnostic accuracy in the Word Dictation task (sensitivity = 1, specificity = .82, AUC = .92) but not in the Picture Word task (sensitivity = .69, specificity = .60, AUC = .76) (Keller-Margulis et al., 2019).

Total Words Written (TWW)

This is a measure that appears in word Dictation, Sentence Writing, and Picture Word tasks. The *validity* of this measure showed significant correlation coefficients ranging from .32 to .48, using the TEWL-2 and WJ-IV as criterion measures (Coker & Ritchey, 2010; Keller-Margulis et al., 2019). Regarding "reliability", IRR indices ranged from .86 to 1 (Coker & Ritchey, 2010; Keller-Margulis et al., 2019), and the alternate form obtained a score of .76 (Coker & Ritchey, 2010). In terms of "growth sensitivity", the estimated slope was 1.50 (*p* = .0001); however, this analysis was conducted with a combined sample of kindergarten and first-grade students (Coker & Ritchey, 2010). Concerning "diagnostic accuracy", the TWW was not precise in either the Word Dictation task (sensitivity = .75, specificity = .58, AUC = .75) or the Picture Word task (sensitivity = .69, specificity = .64, AUC = .74) (Keller-Margulis et al., 2019).

Table 4. CBM-W into the Tasks

CBM - W	Tasks					Characteristics				
	Sentence Writing (Coker & Ritchey, 2010)	Letter Name Knowledge (Justi et al., 2020)	Word Dictation (Keller-Margulis et al., 2019; Ritchey et al., 2010)	Picture Word (Keller-Margulis et al., 2019)	Alphabetic Writing Fluency (Puranik et al., 2017)	Total	Validity	Reliability	Sensitivity to growth	Diagnostic accuracy
TWW	+		+	+		3	Yes	Yes	~	
CSW	+		+ +	+		4	Yes	Yes	Yes	Yes [a]
CWS	+			+		2	Yes	Yes	~	
CL		+	+		+	3	Yes	Yes [a]	Yes [a]	
TLW			+	+		2	Yes	Yes		
CLS			+ +	+		3	Yes	Yes	Yes	Yes [a]
PC			+			1	Yes	Yes	Yes	

Note. + = Number of times this measure is studied in each of the tasks; yes = meets this characteristic; ~ = measure studied with combined kindergarten and 1st grade samples; [a] only fulfills it in word level tasks; TWW = Total Words Written; CSW = Correct Spelled Words; CWS = Correct Word Sequences; CL = Correct Letters; TLW = Total Letters Written; CLS = Correct Letter Sequences; PC = Phonological Coding.

Correct Letters (CL)

It is a measure that appears in subword-level tasks such as Letter-Named Knowledge and Alphabetic Writing Fluency, as well as in word-level tasks such as Word Dictation. In the first two tasks, this measure functions as a count of correctly identified or written letters, with 1 point awarded for each correct letter (Justi et al., 2020; Puranik et al., 2017). However, in the Word Dictation task this measure is treated as a correct letter-sound measure, where 1 point is given for each correct letter or homophone corresponding to the target letter's sound in the dictated words. For example, if the word /kat/ is dictated and the student writes /kat/, they receive 3 points for 3 correct letters. Similarly, if they write /cat/, they also receive 3 points since the sounds /k/ and /c/ are homophones and thus considered correct (Ritchey et al., 2010). Thus, at the subword level, the "validity" of this measure ranges from .21 to .64, using criterion measures such as the TEWL-3, WJ-III, and other tasks such as Picture Word or Essay Prompt (Puranik et al., 2017). At the word level, the correlation coefficients with the TEWL-2 criterion range from .36 to .59 (Ritchey et al., 2010). Regarding "reliability" indices, no reliability data are reported at the subword level, but at the word level, the IRRs reach 95.1% agreement (Ritchey et al., 2010). "Growth sensitivity" was only analyzed at the word level, with an estimated slope of 2.11 ($p = .0001$) (Ritchey et al., 2010). "Diagnostic accuracy" was reported in only one study at the subword level, where sensitivity and specificity scores exceeded .80, but the area under the curve (AUC) ranged between .50 and .60 in both versions of the task (Justi et al., 2020).

Correct Letter Sequences (CLS)

This appears in tasks such as word dictation and picture words. For the "validity" of this measure, the TEWL-2 and WJ-IV were used as criterion measures, and CLS achieved significant correlation coefficients ranging from .33 to .68 (Keller-Margulis et al., 2019; Ritchey et al., 2010). The "reliability" indices ranged from .86 to 1 (Keller-Margulis et al., 2019) and showed 98.7% agreement between the raters (Ritchey et al., 2010). In terms of "growth sensitivity", the estimated slope was 2.94 ($p = .0001$) (Ritchey et al., 2010). Concerning "diagnostic accuracy", CLS demonstrated better indices in the Word Dictation task (sensitivity = .94, specificity = .81, AUC = .93) than in the Picture Word task (sensitivity = .69, specificity = .76, AUC = .83) (Keller-Margulis et al., 2019).

Correct Word Sequences (CWS)

It appears in tasks requiring sentence generation, such as sentence writing or picture writing. The "validity" of this measure was calculated using the TEWL-2 and WJ-IV as criterion measures, yielding values between .20 and .60 (Coker & Ritchey, 2010; Keller-Margulis et al., 2019). Regarding "reliability", IRR indices ranged from .86 to 1 (Coker & Ritchey, 2010; Keller-Margulis et al., 2019), and the alternate form obtained a score of .78 (Coker & Ritchey, 2010). In terms of "growth sensitivity", the estimated slope value was 1.10 ($p = .0001$); however, this calculation used a combined sample of kindergarten and first-grade students (Coker & Ritchey, 2010). Concerning diagnostic accuracy, the indices were acceptable (sensitivity = .81, specificity = .77, AUC = .83) (Keller-Margulis et al., 2019).

Total Letters Written (TLW)

This appears in the Word Dictation and Picture Word tasks. Regarding the "validity" of this measure, correlation coefficients between .41 and .49 were obtained using the WJ-IV as the criterion measure. For "reliability", the IRR indices ranged from .86 to 1. There are no studies reporting "growth sensitivity" for this measure. In terms of "diagnostic accuracy", TLW did not achieve sufficient scores in either the Word Dictation task (sensitivity = .75, specificity = .71, AUC = .78) or the Picture Word task (sensitivity = .75, specificity = .53, AUC = .69) (Keller-Margulis et al., 2019).

Phonological Coding (PC)

It was only used in the word dictation task by Ritchey et al. (2010). In this measure, each word written by students in this task was scored. Each word was scored from 0 to 6, where a lower score indicates little to no relationship between the letters and sounds of the target word, and a higher score indicates a more accurate representation of the sounds, up to correct spelling. The "validity" of this measure showed correlation coefficients between .35 and .58 with the criterion measure, which was the TEWL-2. For the "reliability" of this measure, the percentage of agreement between the raters was 97%. In terms of "growth sensitivity", this measure obtained an estimated slope value of 3.36 ($p = .0001$). There is no reported information on the "diagnostic accuracy" of this measure.

Study Quality Results

Using the CRF-QS, all six articles included in the review underwent a comprehensive quality assessment. Among them, three studies demonstrated excellent methodological rigor, while two were deemed to have good quality (see Table 3). All reviewed articles received favorable scores in the following sections: study purpose, literature review, study design, detailed sample description, results reporting, and conclusions. However, the items where most absences were observed were justifying the sample size or reporting drop-outs. Nevertheless, there were no studies with low methodological quality among the articles included in this review.

Discussion

The aim of the present study was to summarize the psychometric features of the CBM-W in kindergartens. Our review aligns with previous studies that highlight the limited scientific literature on CBM tools in writing (McMaster et al., 2011), particularly in early grades (McMaster & Espin, 2007). Of the 311 articles identified for review, only 5 met the eligibility criteria for the purpose of the study. Furthermore, the reviewed articles display considerable heterogeneity in terms of the tasks employed and the measurement scores observed. All tasks and measures were focused on assessing transcription skills, which are potent predictors of early writing proficiency (Berninger et al., 2002; Coker, 2006; Graham et al., 1997; Kim, et al., 2011; Kirby et al., 2021; Tortorelli et al., 2022). We observed a greater prevalence of production-dependent scoring than of production-independent scoring or precision scoring. These production-dependent scores encompass levels ranging from sentence and discourse to word or subword.

Overall, all CBM writing measures were valid in at least one of their versions compared to criterion measures. Furthermore, all measures reviewed in this study reported reliability measures; however, the CL measure only presents reliability indices within word-level tasks but not in subword-level tasks. Regarding growth sensitivity, these data are available for 6 out of the 7 CBM-W measures studied in this review. However, it is worth noting that the growth sensitivity analyses of the TWW and CWS measures should be interpreted with caution, as the authors combined samples from kindergarten and first grade for this analysis. Therefore, it is difficult to determine whether sensitivity to changes in these measures could appear only in kindergarten,

only in first grade, or in both grades. On the other hand, the growth sensitivity of the CL measure was only studied in word-level tasks, such as word dictation, and not in subword-level tasks, such as the alphabetic writing fluency or letter name knowledge task. Diagnostic accuracy was the least reported characteristic in the studies. Only the CSW and CLS measures showed highly precise diagnostic accuracy, meeting the criteria of obtaining scores close to 1 for sensitivity and specificity and scores above .80 for the area under the curve (AUC) (Mandrekar, 2010). Overall, the measures with the most information available based on their validity, reliability, growth sensitivity, and diagnostic accuracy are the CSW (at the word level) and CLS measures. The next most robust measures are those that meet the criteria for validity, reliability, and sensitivity to growth, but for which we lack information on their diagnostic accuracy. These measures are PC and CL, the latter being applicable only in word-level tasks.

Based on these data, it can be inferred that, as of now, productivity measures reliant on transcription skills, particularly those predicated on spelling metrics rather than handwriting metrics, exhibit validity, reliability, sensitivity to growth, and accuracy during early developmental stages for assessing potential writing difficulties. Furthermore, some authors have reported the floor effect of kindergarten scores when production measures are at the sentence level or when fluency is included (Coker & Ritchey, 2010; Puranik et al., 2017; Ritchey et al., 2010). One possible explanation for this might be that young children begin by establishing the foundations of writing through transcription skills (Ritchey et al., 2015), focusing their cognitive resources on retrieval of the alphabetic code or phoneme-grapheme correspondences (Berninger & Swanson, 1994; Ehri, 1986; McCutchen, 2006). Although we have not found studies that focus on precision measures, it could be a future goal for researchers to add them to investigations given that CSW and CLS measures appear to be robust measures in these populations, which would allow for the calculation of %CSW or %CLS in kindergarten.

The paucity of studies focused on early CBM-W in transparent orthographies is important. Our findings reveal a predominance of such studies conducted with English-speaking students. A notable exception is a study by Justi et al. (2020) involving Portuguese-speaking students. These findings emphasize the imperative to account for the unique features of different orthographic systems within which students learn to write. This is because learners face distinct challenges depending on the type of orthographic system (Arfé et al., 2016; Defior et al., 2009; O'Brien et al., 2020). Therefore, CBM in writing must be customized in line with the specific demands and attributes of each orthographic system. It is essential to improve the scientific evidence in transparent orthographies to establish a solid foundation for evaluating and monitoring writing development in such contexts.

The main limitation of the present study was the scarcity of articles examining the technical characteristics of CBM-W tools designed for kindergarten students. Another limitation stems from the scarcity of comparative data. This is evident in how many measures did not present uniform data across studies. For example, the CSW measure used the TEWL-2 as a criterion measure in some studies to demonstrate its validity, while other studies used the WJ-IV as the criterion measure. Another example is the inconsistency in evaluating measures across tasks with similar demands. This is reflected in the CL measure, where the tasks used to assess it had different demands (writing alphabet letters, naming letters, or writing dictated words). Consequently, it is difficult to establish uniformity in the results, which complicates the generalization of the findings. Other authors have also concluded that depending on the task used, the validity and predictive utility indices of a CBM writing measure may vary (Gil & Jiménez, 2019; Keller-Margulis et al., 2015; McMaster & Espin, 2007). These two constraints present obstacles to conducting a comprehensive analysis of the review. Having more research that addresses CBM writing measures

evaluated in the kindergarten population would resolve many of these limitations by providing a more consistent and comprehensive dataset. This would enable more accurate comparisons across studies, improve the generalizability of findings, and facilitate the development of standardized assessment protocols. Furthermore, it would allow for a better understanding of how different tasks and criterion measures influence the validity and reliability of CBM writing measures, ultimately leading to more effective assessment tools for early writing difficulties.

Conclusion

In conclusion, our study synthesized the psychometric properties of CBM-W tools in kindergarten, revealing a paucity of research in this domain, particularly within transparent orthographic systems. While the reviewed studies demonstrate validity, reliability, sensitivity to growth, and diagnostic accuracy across various measures, there remains variability in task selection and scoring methods. Notably, measures such as Correctly Spelled Words (CSW) and Correct Letter Sequences (CLS) stand out for their robust psychometric profiles, exhibiting high diagnostic accuracy. However, challenges persist, including a dearth of studies in transparent orthographies beyond English-speaking contexts and limitations in task standardization and analysis uniformity. Enhancing the evidence base in transparent orthographies and refining task selection and analysis protocols are imperative for advancing the assessment and monitoring of early writing development. Despite the limitations inherent in the current literature, CBM-W tools, particularly those focusing on transcription skills, show promise in detecting and evaluating writing difficulties in young learners, providing valuable insights into their developmental trajectories. Future research should aim to expand the evidence base for CBM-W in kindergarten to improve cross-study comparisons and develop standardized assessment protocols. This would enhance the understanding of the validity and reliability of CBM-W measures and lead to more effective assessment tools for early writing difficulties. Having assessment measures available for early childhood education could enhance the early detection of writing difficulties by identifying students at potential risk of writing challenges, enabling early interventions, and ultimately improving their long-term outcomes.

Conflict of Interest

The authors of this article declare no conflict of interest.

References

American Psychiatric Association. (2013). *DSM-5. Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.

Ardoin, S. P. (2006). The response in response to intervention: Evaluating the utility of assessing maintenance of intervention effects. *Psychology in the Schools, 43*(6), 713-725. https://doi.org/10.1002/PITS.20181

Arfé, B., Dockrell, J. E., & De Bernardi, B. (2016). The effect of language specific factors on early written composition: The role of spelling, oral language and text generation skills in a shallow orthography. *Reading and Writing, 29*(3), 501-527. https://doi.org/10.1007/s11145-015-9617-5

Arrimada, M., Fidalgo, R., & Torrance, M. (2022). Evaluación de la composición escrita: una revisión de medidas e instrumentos. In M. Arrimada, R. Fidalgo, & M. Torrance (Coords.), *¿Cómo prevenir las dificultades de aprendizaje en la escritura?: la respuesta a la intervención como marco de referencia* (pp. 149-179). Pirámide. https://dialnet.unirioja.es/servlet/articulo?codigo=8861198

Arrimada, M., Torrance, M., & Fidalgo, R. (2020). El modelo de respuesta a la intervención en escritura: revisión de medidas de evaluación y prácticas instruccionales. *Papeles Del Psicólogo, 41*(1), 54-65. https://doi.org/10.23923/PAP.PSICOL2020.2918

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater reliability in systematic review methodology: Exploring variation in coder

decision-making. *Sociological Methods & Research, 50*(2), 837-865. https://doi.org/10.1177/0049124118799372

Berninger, V. W. (2000). Development of language by hand and its connections with language by ear, mouth, and eye. *Topics in Language Disorders, 20*(4), 65-84. https://doi.org/10.1097/00011363-200020040-00007

Berninger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading: Connections between language by hand and language by eye. *Journal of Learning Disabilities, 35*(1), 39-56. https://doi.org/10.1177/002221940203500104

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. L. Swanson, K. R. Harris, & S. Graham (Eds.). *Handbook of learning disabilities* (pp. 345-363). The Guilford Press. https://psycnet.apa.org/record/2003-02238-020

Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008). Writing problems in developmental dyslexia: Under-recognized and under-treated. *Journal of School Psychology, 46*(1), 1-21. https://doi.org/10.1016/J.JSP.2006.11.008

Berninger, V. W., & Swanson, H. L. (1994). Modifying Hayes and Flower's model of skilled writing to explain beginning and developing writing. In E. C. Butterfield & J. S. Carlson (Eds.), *Advances in cognition and educational practice. Children's writing: Toward a process theory of the development of skilled writing* (Vol. 2, pp. 57-81). JAI Press.

Berninger, V. W., & Winn, W. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 96–114).

Blanco Pérez, M., & Bermejo, V. (2009). El efecto Mateo en niños con dificultades específicas de aprendizaje de las matemáticas. *Escritos de Psicología/Psychological Writings, 3*(1), 30-36. https://doi.org/10.24310/ESPSIESCPSI.V3I1.13332

Bosch, R., Pagerols, M., Rivas, C., Sixto, L., Bricollé, L., Español-Martín, G., Prat, R., Ramos-Quiroga, J. A., & Casas, M. (2021). Neurodevelopmental disorders among Spanish school-age children: Prevalence and sociodemographic correlates. *Psychological Medicine*, 1-11. https://doi.org/10.1017/S0033291720005115

Cappa, C., Giulivi, S., Schilirò, A., Bastiani, L., Muzio, C., & Meloni, F. (2015). A screening on specific learning disorders in an Italian speaking high genetic homogeneity area. *Research in Developmental Disabilities, 45-46*, 329-342. https://doi.org/10.1016/J.RIDD.2015.07.011

Coker, D. (2006). Impact of first-grade factors on the growth and outcomes of urban schoolchildren's primary-grade writing. *Journal of Educational Psychology, 98*(3), 471-488. https://doi.org/10.1037/0022-0663.98.3.471

Coker, D. L., & Ritchey, K. D. (2010). Curriculum-based measurement of writing in kindergarten and first grade: An investigation of production and qualitative scores. *Exceptional Children, 76*(2), 175-193. https://doi.org/10.1177/001440291007600203

Defior, S., Jiménez-Fernández, G., & Serrano, F. (2009). Complexity and lexicality effects on the acquisition of Spanish spelling. *Learning and Instruction, 19*(1), 55-65. https://doi.org/10.1016/J.LEARNINSTRUC.2008.01.005

Defior, S., Serrano, F., & Gutiérrez, N. (2015). *Dificultades específicas de aprendizaje*. Editorial Síntesis.

de León, S. C., Jiménez, J. E., García, E., & Gutiérrez, N. (2021a). Identification of Spanish third graders at risk of math problems: Usefulness of number sense based screening measures. *Psychology in the Schools, 58*(7), 1416-1431. https://doi.org/10.1002/PITS.22525

de León, S. C., Jiménez, J. E., García, E., Gutiérrez, N., & Gil, V. (2021b). Universal screening in mathematics for Spanish students in first grade. *Learning Disability Quarterly, 44*(2), 123-135. https://doi.org/10.1177/0731948720903273

Deno, S., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. (1982). *The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study* (Research Rpt. 87). Institute for Research on Learning Disabilities.

Ehri, L. C. (1986). Sources of difficulty in learning to spell and read. *Advances in Developmental & Behavioral Pediatrics, 7*, 121-195.

Faber, I. R., Bustin, P. M. J., Oosterveld, F. G. J., Elferink-Gemser, M. T., & Nijhuis-Van der Sanden, M. W. G. (2015). Assessing personal talent determinants in young racquet sport players: A systematic review. *Journal of Sports Sciences, 34*(5), 395-410. https://doi.org/10.1080/02640414.2015.1061201

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365-387. https://doi.org/10.2307/356600

Fortes, I. S., Paula, C. S., Oliveira, M. C., Bordin, I. A., de Jesus Mari, J., & Rohde, L. A. (2016). A cross-sectional study to assess the prevalence of DSM-5 specific learning disorders in representative school samples from the second to sixth grade in Brazil. *European Child and Adolescent Psychiatry, 25*(2), 195-207. https://doi.org/10.1007/S00787-015-0708-2/TABLES/4

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93-99. https://doi.org/10.1598/RRQ.41.1.4

Fuchs, L. S., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Exceptional Children, 39*(5), 14-20. https://doi.org/10.1177/004005990703900503

Gansle, K. A., Noell, G. H., VanDerHeyden, A. M., Naquin, G. M., & Slider, N. J. (2002). Moving beyond total words written: The reliability, criterion validity, and time cost of alternate measures for curriculum-based measurement in writing. *School Psychology Review, 31*(4), 477-497. https://doi.org/10.1080/02796015.2002.12086169

Gansle, K. A., VanDerHeyden, A. M., Noell, G. H., Resetar, J. L., & Williams, K. L. (2006). The technical adequacy of curriculum-based and rating-based measures of written expression for elementary school students. *School Psychology Review, 35*(3), 435-450. https://doi.org/10.1080/02796015.2006.12087977

Gil, V., & Jiménez, J.E. (2019). Evaluación basada en el currículo dentro del modelo RtI: elemento crítico para su implementación. In J. E. Jiménez (Coord.), *Modelo de respuesta a la intervención: un enfoque preventivo para el abordaje de las dificultades específicas de aprendizaje* (pp. 85-105). Ediciones Pirámide.

Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology, 89*(1), 170-182. https://doi.org/10.1037/0022-0663.89.1.170

Graham, S., Collins, A. A., & Rigby-Wills, H. (2017). Writing characteristics of students with learning disabilities and typically achieving peers: A meta-analysis. *Exceptional Children, 83*(2), 199-218. https://doi.org/10.1177/0014402916664070

Graham, S., Gillespie, A., & Mckeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing, 26*(1), 1-15. https://doi.org/10.1007/s11145-012-9395-2

Graham, S., & Hall, T. E. (2016). Writing and writing difficulties from primary grades to college. *Learning Disability Quarterly, 39*(1), 3-4. https://doi.org/10.1177/0731948715592154

Graham, S., & Harris, K. R. (2005). Improving the writing performance of young struggling writers: Theoretical and programmatic research from the center on accelerating student learning. *The Journal of Special Education, 39*(1), 19-33. https://doi.org/10.1177/00224669050390010301

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445-476. https://doi.org/10.1037/0022-0663.99.3.445

Gutiérrez, N., Jiménez, J. E., & de León, S. C. (2021). Reading curriculum-based measures for Universal Screening in monolingual Spanish first graders. *Early Education and Development, 33*(6), 1036-1060. https://doi.org/10.1080/10409289.2021.1935537

Hampton, D. D., Lembke, E. S., Lee, Y. S., Pappas, S., Chiong, C., & Ginsburg, H. P. (2012). Technical adequacy of early numeracy curriculum-based progress monitoring measures for kindergarten and first-grade students. *Assessment for Effective Intervention, 37*(2), 118-126. https://doi.org/10.1177/1534508411414151

Hooper, S. R., Costa, L.-J., Mcbee, M., Anderson, K. L., Yerby, D. C., Knuth, S. B., Childress, A., Hooper, S. R., Anderson, Á. K. L., Yerby, Á. D. C., Costa, L.-J., Knuth, Á. S. B., Childress, Á. A., Mcbee, M., & Graham, F. P. (2011). Concurrent and longitudinal neuropsychological contributors to written language expression in first and second grade students. *Reading and Writing, 24*(2), 221-252. https://doi.org/10.1007/s11145-010-9263-x

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. www.kansasmtss.org

Jewell, J., & Malecki, C. K. (2005). The utility of CBM written language indices: An investigation of production dependent, production-independent, and accurate-production scores. *School Psychology Review, 34*(1), 27-44. https://doi.org/10.1080/02796015.2005.12086273

Jiménez, J. E. (2018). *Early Grade Writting Assessment: A report on development of an instrument.* United Nations Educational, Scientific, and Cultural Organization.

Jiménez, J. E., De León, S. C., & Gutiérrez, N. (2021). Piloting the response to intervention model in the Canary Islands: Prevention of reading and math learning disabilities. *Spanish Journal of Psychology, 24.* https://doi.org/https://doi.org/10.1017/SJP.2021.25

Jiménez, J. E., & García De La Cadena, C. (2007). Learning disabilities in Guatemala and Spain: A cross-national study of the prevalence and cognitive processes associated with reading and spelling disabilities. *Learning Disabilities Research & Practice, 22*(3), 161-169. https://doi.org/https://doi.org/10.1111/j.1540-5826.2007.00247.x

Johnson, E., Mellard, D. F., Fuchs, D., & McKnight, M. A. (2006). *Responsiveness to Intervention (RTI): How to do it. RTI manual.* National Research Center on Learning Disabilities. www.nrcld.org

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy. A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*(4), 243-255. https://doi.org/10.1037/0022-0663.78.4.243

Justi, C. N. G., Cunha, N., & Justi, F. R. dos R. (2020). Letter-name knowledge: Predicting reading and writing difficulties. *Estudos de Psicologia (Campinas), 37*(3). https://doi.org/10.1590/1982-0275202037e180173

Katusic, S. K., Colligan, R. C., Weaver, A. L., & Barbaresi, W. J. (2009). The forgotten learning disability: Epidemiology of written-language

disorder in a population-based birth cohort (1976-1982), Rochester, Minnesota. *Pediatrics, 123*(5), 1306-1313. https://doi.org/10.1542/peds.2008-2098

Keller-Margulis, M. A., Ochs, S., Reid, E. K., Faith, E. L., & Schanding, Jr. G. T. (2019). Validity and diagnostic accuracy of early written expression screeners in kindergarten. *Journal of Psychoeducational Assessment, 37*(5), 539-552. https://doi.org/10.1177/0734282918769978

Kent, S., Wanzek, J., Petscher, Y., al Otaiba, S., Kim, Y.-S., Kent, S., Wanzek, J., Kim, Y.-S., Petscher, Y., & al Otaiba, S. (2014). Writing fluency and quality in kindergarten and first grade: The role of attention, reading, transcription, and oral language. *Reading and Writing, 27*, 1163-1188. https://doi.org/10.1007/s11145-013-9480-1

Kim, Y. S., Al Otaiba, S., Puranik, C., Folsom, J. S., Greulich, L., & Wagner, R. K. (2011). Componential skills of beginning writing: An exploratory study. *Learning and Individual Differences, 21*(5), 517-525. https://doi.org/10.1016/J.LINDIF.2011.06.004

Kirby, M. S., Spencer, T. D., & Chen, Y.-J. I. (2021). Oral narrative instruction improves kindergarten writing. *Reading and Writing Quarterly, 37*(6), 574-591. https://doi.org/10.1080/10573569.2021.1879696

Landerl, K., & Moll, K. (2010). Comorbidity of learning disorders: Prevalence and familial transmission. *Journal of Child Psychology and Psychiatry, 51*(3), 287-294. https://doi.org/10.1111/J.1469-7610.2009.02164.X

Landis, J. R, & Koch, G. K. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-74. https://doi.org/10.2307/2529310

Law, M., Stewart, D., Pollock, N., Bosch, J., & Westmorland, M. (1998). *Guidelines for critical review form - quantitative studies.* CanChild, Research Centre of McMaster University.

Lembke, E., Deno, S. L., & Hall, K. (2016). Identifying an indicator of growth in early writing proficiency for elementary school students. *Assessment for Effective Intervention, 28*(34), 23-35. https://doi.org/10.1177/073724770302800304

Leonardi, M. M., Domenico, F., Blasi, D., Savelli, E., & Buono, S. (2021). Reading and spelling disorders in a school-based population screening in Sicily (Italy). *Dyslexia, 27*(4), 452-467. https://doi.org/10.1002/dys.1694

Lonigan, C., Schatschneider, C., & Westberg, L. (2008). Identification of children's skills and abilities linked to later outcomes in reading, writing, and spelling. *Developing early literacy: Report of the National Early Literacy Panel,* 55-106.

MacArthur, C. A., & Philippakos, Z. A. (2013). Self-regulated strategy instruction in developmental writing: A design research project. *Community College Review, 41*(2), 176-195. https://doi.org/10.1177/0091552113484580

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology, 5*(9), 1315-1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

McCutchen, D. (2006). Cognitive factors in the development of children's writing. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 115-130). The Guilford Press.

McMaster, K. L., & Espin, C. (2007). Technical features of curriculum-based measures for beginning writers: A literature review. *The Journal of Special Education, 42*(2), 68-84. https://doi.org/10.1177/00224669070410020301

McMaster, K. L., Ritchey, K. D., & Lembke, E. (2011). Curriculum-based measurement for beginning writers: Recent developments and future directions. *Advances in Learning and Behavioral Disabilities, 24,* 111-148. https://doi.org/10.1108/S0735-004X(2011)0000024008

Moll, K., Kunze, S., Neuhoff, N., Bruder, J., & Schulte-Kö Rne, G. (2014). Specific learning disorder: Prevalence and gender differences. *PLoS ONE, 9*(7), Article 103537. https://doi.org/10.1371/journal.pone.0103537

National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the educational success of children and youth learning English. Promising futures.* The National Academic Press. https://doi.org/10.17226/24677

National Center on Response to Intervention. (2010). *Information brief: Developing an RTI guidance document.* https://files.eric.ed.gov/fulltext/ED578037.pdf

O'Brien, B. A., Habib Mohamed, M. B., Arshad, N. A., & Lim, N. C. (2020). The impact of different writing systems on children's spelling error profiles: Alphabetic, Akshara, and Hanzi cases. *Frontiers in Psychology, 11.* https://doi.org/10.3389/FPSYG.2020.00870

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews, 5*(1), 1-10. https://doi.org/10.1186/S13643-016-0384-4/FIGURES/6

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ, 372,* Article 71. https://doi.org/10.1136/BMJ.N71

Puranik, C. S., & Lonigan, C. J. (2012). Name-writing proficiency, not length of name, is associated with preschool children's emergent literacy skills. *Early Childhood Research Quarterly, 27*(2), 284 -294. https://doi.org/10.1016/J.ECRESQ.2011.09.003

Parker, D. C., Burns, M. K., McMaster, K. L., & Shapiro, E. S. (2012). Extending curriculum-based assessment to early writing. *Learning Disabilities Research & Practice, 27*(1), 33-43. https://doi.org/10.1111/j.1540-5826.2011.00348.x

Parker, R., Tindal, G., & Hasbrouck, J. (2009). Countable indices of writing quality: Their suitability for screening–eligibility decisions. *Exceptionality: A Special Education Journal, 2*(1), 1-17. https://doi.org/10.1080/09362839109524763

Puranik, C. S., Patchan, M. M., Sears, M. M., & McMaster, K. L. (2017). Examining alphabet writing fluency in kindergarten: Exploring the issue of time on task. *Assessment for Effective Intervention, 42*(2), 81-96. https://doi.org/10.1177/1534508416665324

Ritchey, K. D., Coker, D. L., & McCraw, S. B. (2010). A comparison of metrics for scoring beginning spelling. *Assessment for Effective Intervention, 35*(2), 78-88. https://doi.org/10.1177/1534508409336087

Ritchey, K. D., McMaster, K. L., Al Otaiba, S., Puranik, C. S., Kim, Y. S. G., Parker, D. C., & Ortiz, M. (2015). Indicators of fluent writing in beginning writers. In K. D. Cummings & Y. Petscher (Eds.), *The fluency construct: Curriculum-based measurement concepts and applications* (pp. 21-66). https://doi.org/10.1007/978-1-4939-2803-3_2

Rodríguez, C., Jiménez, J. E., & Balade, J. (2024). The impact of oral language and transcription skills on early writing production in kindergarteners: Productivity and quality. *Early Childhood Education Journal,* 1 -11. https://doi.org/10.1007/S10643-024-01670-4

The National Center on Intensive Intervention. (2021, July). *National Center on Intensive Intervention at the American Institutes for Research.* Academic Screening Tools Chart. https://charts.intensiveintervention.org/ascreening

Tolchinsky, L. (2006). The emergence of writing. In L. Tolchisky (Ed.), *Handbook of writing research* (pp. 83-95).

Tortorelli, L. S., Gerde, H. K., Rohloff, R., & Bingham, G. E. (2022). Ready, set, write: Early learning standards for writing in the common core era. *Reading Research Quarterly, 57*(2), 729–752. https://doi.org/10.1002/RRQ.436

United Nations Educational, Scientific, and Cultural Organization (UNESCO, 1977). *UNESCO Thesaurus. Vocabulary Information.* https://vocabularies.unesco.org/browser/thesaurus/en/?clang=en

Vélez Calvo, X. M. (2017). *Análisis de la inclusión educativa a través de indicadores de prevalencia de dificultades de aprendizaje, actitudes del profesorado y condiciones de accesibilidad en los centros de la ciudad de Cuenca (Ecuador).* Universitat de València.

Videen, J., Deno, S., & Marston, D. (1982). *Correct word sequences: A valid indicator of proficiency in written expression* (Rsch. Rpt., No. 84).

VanDerHeyden, A. M., Witt, J. C., Naquin, G., & Noell, G. (2001). The reliability and validity of curriculum-based measurement readiness probes for kindergarten students. *School Psychology Review, 30*(3), 363 -382. https://doi.org/10.1080/02796015.2001.12086121

Von Aster, M., Schweiter, M., & Zulauf, M. W. (2010). Rechenstörungen bei Kindern Vorläufer, Prävalenz und psychische Symptome. *Journal of Child Psychology and Psychiatry, 51*(3), 287-294. https://doi.org/10.1026/0049-8637.39.2.85

Whitaker, D., Berninger, V., Johnston, J., & Lee Swanson, H. (1994). Intraindividual differences in levels of language in intermediate grade writers: Implications for the translating process. *Learning and Individual Differences, 6*(1), 107-130. https://doi.org/10.1016/1041-6080(94)90016-7