Revista Colombiana de Educación N. 101 Año 2026
ISSN 0120-3916 / e-ISSN: 2323-0134
DOI: https://doi.org/10.17227/rce.num101-20380
Universidad Pedagógica Nacional, Colombia

Adaptation and Validation of the IPAL for Screening Reading Skills in Colombian First-Grade Students

Received: December 4, 2023
Evaluated: January 3, 2025
Published: July 1, 2026

Yalov Villadiego Ramirez*
Juan E. Jiménez**
Mayilin Moreno-Torres***

Type: Research article

Abstract

Curriculum-based measures (CBMS) are screening assessments designed to evaluate proficiency in foundational areas such as reading, mathematics, and writing, aligning with established curriculum standards. These educational assessment tools have consistently proven highly effective in promptly identifying students in need of additional support, significantly bolstering their academic progress. This study aimed to analyze the factor structure and diagnostic accuracy of the adaptation of the indicators of basic early reading skills (IPAL) to the Colombian context. The IPAL-C is a curriculum-based measure (CBM) designed for screening reading skills in first-grade students. The ABCDETI test was also used as a criterion to analyze the diagnostic accuracy of the IPAL-C. The results showed that IPAL-C demonstrated remarkable classification capacity, especially in the middle and at the end of the school year. Oral reading fluency and pseudoword reading were identified as significant predictors of risk for reading disabilities.

Keywords:

ROC curve; curriculum-based measurement; screening; progress monitoring; reading

* PhD, Universidad del Norte. yvilladiego@uninorte.edu.co
** PhD, Instituto Universitario de Neurociencia (IUNE), Universidad de La Laguna. ejimenez@ull.edu.es
*** PhD, Universidad del Norte. mamoreno@uninorte.edu.co

Adaptação e validação do IPAL para o rastreamento de habilidades de leitura em estudantes colombianos do primeiro ano do ensino fundamental

Resumo

As Medidas Baseadas no Currículo (MBC) são avaliações de triagem destinadas a medir o domínio de conhecimentos e habilidades em áreas fundamentais como leitura, escrita e matemática, de acordo com os padrões curriculares estabelecidos. Esses instrumentos de avaliação educacional demonstraram ser altamente eficazes na identificação precoce de estudantes que necessitam de apoio adicional, contribuindo significativamente para seu progresso acadêmico. Este estudo teve como objetivo analisar a estrutura fatorial e a precisão diagnóstica da adaptação do instrumento Indicadores de Progresso de Aprendizagem em Leitura (IPAL) ao contexto colombiano. Trata-se de uma ferramenta de triagem universal destinada a detectar estudantes em risco de apresentar dificuldades na leitura. Para avaliar a precisão diagnóstica do IPAL-C, utilizou-se o teste ABCDETI como critério, e os resultados indicaram que o IPAL-C possui uma notável capacidade de classificação, especialmente no meio e no final do ano letivo. Além disso, identificou-se que a fluência leitora e a leitura de pseudopalavras são preditores significativos do risco de dificuldades de aprendizagem em leitura.

Palavras-chave

curva ROC; avaliação baseada no currículo; testes de triagem; monitoramento do progresso; leitura

Adaptación y validación del IPAL para el tamizaje de habilidades lectoras en estudiantes colombianos de primer grado

Resumen

Las Medidas Basadas en el Currículo (MBC) constituyen evaluaciones de tamizaje destinadas a medir el dominio de conocimientos y habilidades en áreas fundamentales como lectura, escritura y matemáticas, de acuerdo con los estándares curriculares establecidos. Estos instrumentos de evaluación educativa han demostrado ser altamente efectivos al identificar con prontitud a aquellos estudiantes que requieren apoyo adicional, lo que contribuye significativamente a su progreso académico. Este estudio tuvo como objetivo analizar la estructura factorial y la precisión diagnóstica de la adaptación del instrumento Indicadores de Progreso de Aprendizaje en Lectura (IPAL) al contexto colombiano. Se trata de una herramienta de tamizaje universal destinada a detectar a estudiantes en riesgo de presentar dificultades en la lectura. Para evaluar la precisión diagnóstica del IPAL-C, se utilizó la prueba ABCDETI (Cadavid-Ruiz et al., 2016) como criterio, y los resultados mostraron que el IPAL-C tiene una notable capacidad de clasificación, especialmente a mitad y al final del año escolar. Además, se identificó que la fluidez lectora y la lectura de pseudopalabras son predictores significativos del riesgo de presentar dificultades de aprendizaje en lectura.

Palabras clave

curva ROC; evaluación basada en el currículo; pruebas de tamizaje; monitoreo del progreso; lectura

To cite this article:

Villadiego Ramirez, Y., Jiménez, J. E. y Moreno-Torres, M. (2026). Adaptation and Validation of the IPAL for Screening Reading Skills in Colombian First-Grade Students. *Revista Colombiana de Educación,* (101), e20380, https://doi.org//10.17227/rce.num101-20380

Introduction

Curriculum-based measures (CBMs) are assessment tools designed to measure fundamental skills aligned with the learning objectives established in school curricula (Fuchs & Fuchs, 2011; MacLane, 2008). Their application in educational screening has proven effective in the early detection of difficulties, allowing timely interventions. In this context, the present study focuses on the adaptation and validation of the IPAL-C, a CBM originally developed in Spain (Jiménez & Gutiérrez, 2019), for its implementation with first-grade students in Colombia. In contrast to traditional psychometric measures commonly used in education, some assessments may fail to demonstrate the effectiveness of instruction or intervention (Barnett et al., 2000) and may not be sensitive to changes in the learning process (Salvia et al., 2012). This underscores the need for psychometrically robust formative assessment tools for early identification and intervention. In the educational field, screening assessments allow for the rapid and efficient identification of students at risk of academic difficulties. However, in the Colombian context, there are limitations in the availability of standardized tools aligned with educational standards or learning guidelines. The IPAL-C is an evidence-based tool adapted to Colombian Spanish and aligned with local educational needs. Its implementation facilitates the early detection of reading difficulties in first grade. The use of assessment tools aligned with the school curriculum is essential to ensuring equity in access to education and the early identification of learning difficulties. In this regard, previous studies have emphasized the importance of standardized instruments that allow for the evaluation of academic performance from the early stages of education (Durán-Chiappe & Cruz-Velásquez, 2024). Similarly, research on early education has shown that learning gaps can be mitigated through screening and assessment strategies that effectively identify students at risk and facilitate timely interventions (Cortez-Soto & Moreno-Treviño, 2024). These findings reinforce the relevance of the IPAL-C as a tool that not only monitors students' reading progress but also contributes to the formulation of evidence-based educational policies. These assessments assist educators in evaluating students' levels of knowledge and skills, enabling the identification of potential learning gaps. Additionally, they allow for appropriate interventions or further assessments to address these gaps (UK National Screening Committee, NSC, 2009). As highlighted in the report by Abusamra et al. (2024), the implementation of theoretically grounded and norm-referenced assessment tools is essential for the early detection of literacy difficulties, which in turn activates specific actions for their mitigation in the Ibero-American context.

The IPAL has previously been validated in Spain, where it has demonstrated its utility as a CBM for screening reading difficulties and its sensitivity to learning-related changes (Gutiérrez et al., 2021). On the basis of this empirical foundation, the present study aims to adapt and validate the IPAL-C in the Colombian context by evaluating its factorial structure and diagnostic accuracy in first-grade students. Through this validation, we seek to determine its capacity to reliably identify students at risk of reading difficulties, as well as its alignment with the learning standards or guidelines established by the Colombian Ministry of National Education (MEN). Similarly, in the field of Learning Disabilities (LD) prevention, research has suggested that the earlier the factors preceding the manifestation of a behavioral or academic performance problem are identified and addressed, the greater the likelihood of effectively managing or even mitigating them. Early intervention not only prevents these issues from significantly impacting academic performance but also contributes to students' emotional well-being (Levitt et al., 2007; Pina-Paredes et al., 2020). In this context, research on CBM provides a valuable theoretical and empirical foundation for developing screening assessments in reading, writing, and mathematics (Jiménez, 2019), which are based on the specific objectives and content of curricula established by educational authorities. CBMs are administered at different points throughout the academic year (beginning, middle, and end) to assess student progress, identify areas requiring additional support, and evaluate the effectiveness of implemented educational interventions. By providing objective and measurable data on student performance, these assessments enable educators to make informed decisions regarding teaching and learning. They offer a comprehensive overview of students' competency levels and facilitate the early detection of those who may require additional educational intervention (Anderson et al., 2020; Fuchs et al., 2004; Nese et al., 2011). In the field of reading acquisition, the National Reading Panel (NRP) (2000) emphasized the importance of early identification of Specific Reading Difficulties (SRDs) and literacy-related problems. Without intensive reading support, low reading proficiency can lead to significant learning difficulties in secondary school (Wember & Greisbach, 2018).

Curriculum-Based Assessment and Reading

In the context of research on the validity of curriculum-based measures in reading (CBM-R), an extensive body of scientific literature is available. Research has been conducted in Spanish-speaking countries, particularly in Spain (Gutiérrez et al., 2021). Given the complexity of the reading process, studies have shown that using measures that assess only a single component of reading is not effective in risk detection. Instead, a combination of measures evaluating different components can increase screening accuracy (Compton et al., 2006; O'Connor

& Jenkins, 1999). Moreover, assessments must be sensitive to reading development, the components of reading, and the grade level for which the test is intended. For example, in the first grade, the most effective screening measures combine the following components: fluency in letter naming and sound identification; phoneme segmentation; vocabulary; and word identification fluency (Acha et al., 2023; Compton et al., 2006; Francis et al., 2005; Jenkins & Johnson, 2008; O'Connor & Jenkins, 1999). Indeed, recent studies have demonstrated that children read with expression only after achieving a high level of reading automation (Rodríguez et al., 2024).

In this context, it is crucial to consider, during the design and implementation of a screening test, the necessity of following an explicit, systematic, and transparent process. This process involves evaluating the effectiveness of the screening tool and determining its feasibility (UK National Screening Committee, 2023). A screening assessment must meet multiple quality criteria to be considered satisfactory (Johnson & Marlow, 2006). Some fundamental criteria include the following: a) Cost-effectiveness refers to the simplicity, ease, and brevity of test administration, its low cost of use, accessibility, and adaptability to different languages and cultures (Alcantud et al., 2015). In summary, the test must be affordable and efficient in terms of the time and resources required for administration and result evaluation. In the case of reading, these tests have proven to be highly effective (Fuchs et al., 2004). b) Reliability: The test results must be consistent, meaning that when the test is administered under the same conditions to the same subjects, it should yield similar results each time, ensuring stable outcomes. To evaluate the reliability of CBM, tests such as test–retest reliability, interrater reliability, and internal consistency are used. In the case of screening measures consisting of parallel forms applied at the beginning, middle, and end of the school year, comparability between forms must be established. This is assessed through the parallel-form reliability coefficient, which considers the mean and standard deviation of each form (Glover & Albers, 2007). c) Validity: The test must accurately and reliably measure the variable it aims to assess, supported by solid evidence that it measures what it is intended to measure.

With respect to the validity of screening assessments, three types of validity must be considered: construct validity, content validity, and criterion validity (Sudaryono et al., 2019). Construct validity determines whether an instrument truly measures the theoretical construct or variable of interest. To establish this, the internal and external structure of the instrument must be analyzed. Content validity indicates whether the items comprising a screening instrument reflect the multidimensional nature of what is being assessed. Criterion validity assesses whether an evaluation can predict how well an individual will perform on a particular criterion. There are two types of criterion validity: concurrent validity

and predictive validity (Glover & Albers, 2007). Concurrent validity assesses how well a measure can identify individuals who currently experience difficulties, using a standardized reference test as the criterion (Johnson et al., 2009). Predictive validity, on the other hand, distinguishes between those who are not at risk and those who are likely to experience difficulties in academic performance or behavior, depending on the variable being assessed. Given the characteristics of screening assessments, predictive validity is considered the most important criterion to fulfill (Glover & Albers, 2007). Sensitivity and specificity indices, predictive values, and likelihood ratios are frequently used to evaluate the predictive utility of screening (Bravo-Grau & Cruz, 2015).

A preventive education system requires a screening process that accurately identifies students at risk on the basis of their current performance in the learning area for which the CBM was designed (e.g., mathematics, reading, writing, or sciences) and their grade level. Therefore, diagnostic accuracy is essential for educational screening assessments for the following reasons: 1) Early identification of difficulties: high accuracy in screening assessments enables the early detection of potential difficulties or needs among students, facilitating timely intervention to address these issues. No test classifies 100% of the subjects correctly in terms of risk (sensitivity) and nonrisk (specificity); therefore, the goal is to minimize the number of misclassified cases (Hill & Lewicki, 2007). 2) Personalized intervention: With accurate results, educators can monitor student progress, tailor teaching strategies, and provide specific support to students in need, thereby enhancing their learning and development opportunities (Deno, 2003; Fuchs & Fuchs, 2011; Liebers et al., 2019). 3) Efficient resource allocation: Accurate screening assessments help allocate resources efficiently by identifying students requiring additional attention, thus preventing unnecessary investment in those who do not need it. Hintze et al. (2003) argued that in the context of screening, false positives are generally less problematic; the priority is to detect those at risk. Therefore, high sensitivity in the test is critical. 4) Evaluation of intervention impact: Diagnostic accuracy also allows for a more precise measurement of the effects of educational interventions, which is essential for continuously improving educational programs and policies (Mellard & Johnson, 2008).

Method

Objectives and Hypotheses

This study aims to examine the construct validity of the IPAL-C and determine whether both the overall IPAL-C and its reading-related components can

effectively identify at-risk and non-at-risk students in first grade. The proposed hypotheses include the following: a) The adaptation of the IPAL to the Colombian context (IPAL-C) will exhibit strong construct validity, similar to findings from previous studies conducted with Spanish samples; b) the parallel forms of the IPAL-C will demonstrate adequate diagnostic accuracy in classifying students as at-risk or non-at-risk for SRD; and c) the combined set of IPAL-C measures will show greater diagnostic accuracy than individual measures applied separately.

Participants

The sample for this study consisted of 292 first-grade primary school students (155 boys, 137 girls; age in months, M = 79.8, SD = 6.66) from five schools in the Barranquilla district, Colombia. Barranquilla is the capital of the Atlántico department, located in the Caribbean region of the country, and is divided into five localities: Suroccidente, Suroriente, Norte-Centro Histórico, Metropolitana, and Riomar. To ensure geographical representativeness, one school was selected from each locality, forming a diverse sample in terms of socioeconomic strata and regional characteristics. The sample comprised three public schools and two private schools. With respect to school classification on the basis of academic performance, according to the Colombian Institute for Educational Evaluation (ICFES), schools in Colombia are categorized into five performance levels on the basis of the results of the Saber 11 standardized tests: A+, A, B, C, and D, with A+ being the highest-performing category and D being the lowest. In this study, one school belonged to category A+ (very high performance), two to category A (high performance), one to category B (low performance), and one to category D (poor performance). Complete class groups were included in the study, resulting in a total of 10 evaluated groups, with two groups per school. Students with diagnosed neurodevelopmental disorders or disabilities (i.e., sensory, physical, motor, or intellectual) were excluded from the study. The selected schools represent diverse socioeconomic and cultural contexts. Consequently, parents' educational levels vary, reflecting the heterogeneity of educational backgrounds and available resources in Barranquilla.

Instruments

Indicadores de Progreso de Aprendizaje en Lectura (IPAL) (Jiménez & Gutiérrez, 2019). The adaptation of this instrument to the Colombian context involved not only linguistic adjustments but also the alignment of the IPAL with the first-grade educational standards or learning guidelines proposed by the Colombian Ministry of National Education (MEN) regarding the reading process. This adaptation, like its original version, includes three parallel assessment forms, Form A (beginning),

Form B (middle), and Form C (end), which are administered at three different points throughout the academic year (April–July–November). The IPAL consists of an instruction manual, a student booklet, and an examiner's scoring template, and its administration requires a stopwatch. This CBM-R assesses the skills identified by the National Reading Panel (2000) as the main components of reading: alphabetic knowledge, vocabulary, phonemic awareness, comprehension, and fluency. In addition to these five components, the IPAL also includes functional knowledge of written language, which, according to the National Early Literacy Panel (2009), has been shown to be one of the best predictors of reading comprehension in early education. The tasks included in the IPAL are as follows: a) Letter name fluency (CNL): Children name as many uppercase and lowercase letters as possible within one minute. The score is based on the number of correctly identified letters. b) Letter sound fluency (CSL): Students identify letters and their corresponding sounds within one minute. The score is based on the number of correctly identified sounds. c) Phonological Awareness–Segmentation (CFS): Students segment nonsense words into phonemes within one minute, with progressively increasing difficulty. The score is based on the number of correctly identified phonemes. d) Functional knowledge of written language (CLE): This task assesses basic knowledge of print through six questions about the structure and use of textbooks. The score is based on the number of correct responses. e) Pseudoword reading (LP): Students read nonsense words within one minute to assess their grapheme-phoneme correspondence skills. The score is based on the number of correctly read words. f) Cloze Task (TM): Students complete incomplete sentences by selecting the correct option from three alternatives within five minutes. The score is based on the number of correct responses. g) Oral reading fluency (FLO): Students read a short text aloud for one minute, and both speed and accuracy are measured. The score is based on the number of correctly read words.

Test ABCDETI (Cadavid-Ruiz et al., 2016). The ABCDETI test was used as a criterion measure to determine students' risk status for SRD in first grade. This test, adapted and validated in Colombia for kindergarten, first, and second grades, assesses eight core reading-related skills through a computerized program, as identified by the authors and documented in the National Reading Panel report (2000). The test was administered alongside the end-of-year IPAL-C assessment (Form C) in November. For first-grade primary school students, the ABCDETI evaluates the following skills: alphabetic principles, phonological awareness, text types, reading fluency and accuracy, reading comprehension, and oral comprehension.

Procedure

The IPAL tool, originally developed for schoolchildren in the Canary Islands (Spain), was reviewed and adapted to the Colombian context (IPAL-C) through a process that included (a) adjustments to vocabulary and linguistic structures to ensure alignment with the Colombian variety of Spanish, (b) alignment with the educational standards or learning guidelines for first-grade language established by the Colombian Ministry of National Education (MEN), and (c) evaluation of the conceptual and psychometric equivalence of the items in relation to the original version. This study adopted a longitudinal design in which the IPAL-C was administered at three points during the academic year to monitor progress in reading skills and assess its diagnostic accuracy in identifying at-risk students. Form A was administered after allowing a period for school adaptation and progress, approximately one and a half months after the beginning of the school year. Forms B and C of the IPAL-C were administered three months apart. The ABCDeti criterion test was administered alongside Form C of the IPAL-C in November to obtain concurrent validity data. Each student's response for every measure was recorded on the examiner's evaluation sheet for all three forms of the IPAL-C and the ABCDeti test. All assessments were administered in a standardized manner by evaluators who had been previously trained in the IPAL-C administration protocol.

Data analysis

The internal structure and construct validity of the IPAL-C were evaluated through Confirmatory Factor Analysis (CFA) at each of the three measurement points. For each CFA, the error term variances were freely estimated, and one structural coefficient in the latent variable or factor was fixed to one. The estimation method used was Robust Maximum Likelihood (MLR), and reliability indices were calculated via Cronbach's alpha (α) and hierarchical omega (Ω) coefficients. Model fit was assessed via absolute fit indices such as chi-square and the Root Mean Square Error of Approximation (RMSEA), and incremental fit indices, including the Tucker-Lewis Index (TLI) and the comparative fit index (CFI), were employed to compare residual differences between the sample covariance matrix and the hypothesized model. This analysis determined whether the tasks comprising the IPAL-C adequately reflected a unidimensional theoretical model of reading competence, in which the assessed skills were associated with a single latent reading factor.

To determine the diagnostic accuracy of the IPAL-C in identifying at-risk students, a Receiver Operating Characteristic (ROC) curve analysis was conducted, and the area under the curve (AUC) along with the sensitivity and

specificity indices were calculated. This method established the instrument's ability to correctly classify students at risk for reading difficulties. The ROC curve is a graphical representation used to evaluate an instrument's ability to distinguish between groups (e.g., individuals at risk and not at risk). The curve illustrates the relationship between sensitivity (true positives) and 1-specificity (false positives) at different cutoff points. The AUC measures the instrument's overall accuracy, where a value of 1 indicates perfect discrimination, and a value of 0.5 suggests no better classification than chance. A specific percentile was initially established via the criterion test administered alongside the screening assessment (Johnson et al., 2009). Accordingly, to classify risk status, the overall scores of both the ABCDETI criterion test and the IPAL-C were calculated. The total ABCDETI score was converted into a dichotomous variable (risk vs. nonrisk), with the 25th percentile used as the threshold for identifying at-risk students. The AUC allowed for testing the classification model's performance in distinguishing students at risk for SRD from those not at risk. The classification criteria were based on the Youden index (J) to select the optimal cutoff point for risk classification. The J index is a metric used in sensitivity and specificity analysis to determine the most effective cutoff point for an instrument. It is calculated as J = Sensitivity + Specificity - 1, with its maximum value indicating the threshold that optimally balances both sensitivity and specificity. These measures determine the quality and accuracy of the screening assessment in terms of sensitivity and specificity.

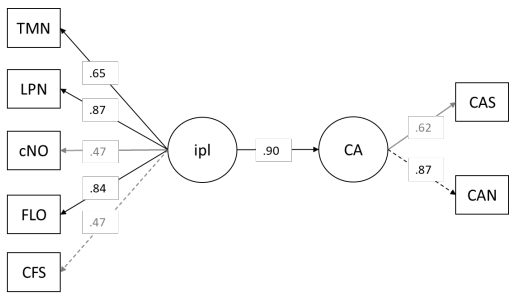
Sensitivity refers to the percentage of students correctly identified as at risk, whereas specificity is the percentage of students correctly identified as not at risk. The screening test results classify students into four groups: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) (National Center on Response to Intervention, 2013). Once sensitivity and specificity were determined, additional indices were calculated, including the predictive value (PV), which estimates the probability that the classification made by the instrument is accurate. Additionally, the positive predictive value (PPV) and negative predictive value (NPV) were evaluated to determine the likelihood that students identified as at risk or not at risk genuinely belonged to their respective categories (Donis, 2012). The likelihood ratio (LR), also referred to as predictive efficiency, was used to determine the precision of the screening assessment by indicating the probability (prevalence) of developing the condition. This measure assesses whether test results increase or decrease the likelihood of being at risk for SRD. Two types of likelihood ratios were considered: the positive likelihood ratio (PLR) and the negative likelihood ratio (NLR) (Altman & Bland, 1994; Donis, 2012).

Results

The factorial structure of each parallel form for the first grade was analyzed via confirmatory factor analysis (CFA). On the basis of this analysis, this study proposes a hypothetical model of reading competence, grounded in empirical data obtained from students' responses via the IPAL-C instrument. Under this model and in light of the theoretical review, the presence of a single first-order factor related to reading competence is supported. Therefore, all tasks included in the IPAL-C (such as CNL, CSL, CFS, CLE, LP, TM, and FLO) across the three assessment points in the academic year (April, July, and November) are linked to this latent factor. This model reflects a unidimensional theoretical conception of reading competence.

In Form A, the χ^2 value obtained was 17.234 (p = .189), suggesting that the model and the data fit well. Analyzing the goodness-of-fit indices obtained during model estimation, the results indicate a good fit to the data (SRMR = 0.031; χ^2 = 17.234; df = 13; p = .189; NFI = .975; CFI = .994; NNFI = .990; MFI = .991; RMSEA = .036). The findings indicate good reliability for Form A of the IPAL-C within the confirmatory factorial model, demonstrating adequate internal consistency (α = .71, ω = .85). These values suggest an acceptable level of internal consistency in the measurement conducted within the framework of the confirmatory factorial model (see Figure 1).

Figure 1.
Measurement Model for Form A: Confirmatory Factor Analysis



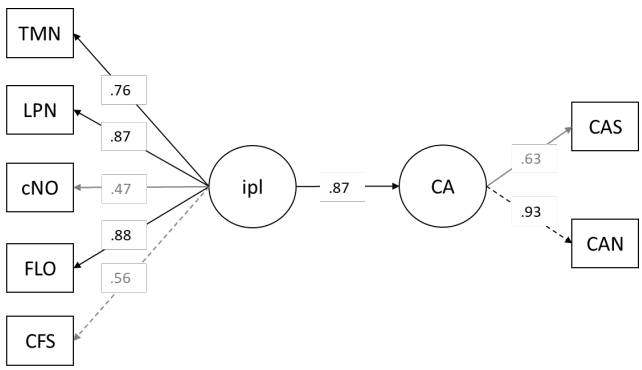
Note. IPL = IPAL-C (April); TMA = Cloze Task (April); LPA = Pseudoword Reading (April); CNO = Functional Knowledge of Written Language (April); FLO = Oral Fluency (April); CFS = Phonological Awareness (April); CA = Alphabetic Knowledge (April); CAS = Letter Sound Knowledge (April); CAN = Letter Name Knowledge (April). Source: Own elaboration.

Additional Note: The confirmatory factor analysis (CFA) for Form A demonstrates an adequate fit between the model and the data, supported by a χ^2

value of 17.234 (p = .189), indicating that the discrepancies between the model and the data are not statistically significant. Goodness-of-fit indices, such as SRMR = .031, CFI = .994, and RMSEA = .036, fall within the recommended thresholds, further reinforcing the model's validity. Additionally, the values of NFI = .975 and NNFI = .990 confirm that the model is appropriate for the data. The reliability of Form A is also adequate, with α = .71 and ω = .85, suggesting satisfactory internal consistency in the measurements obtained. These results indicate that the Form A model has good fit and adequate reliability, supporting its use in confirmatory factor analysis.

In Form B, a χ^2 value of 17.909 (p = .161) was obtained, indicating a reasonable fit between the model and the data. Evaluating the goodness-of-fit indices, the model demonstrates an adequate fit to the empirical data (χ^2/df = 1.377, SRMR = .025, NFI = .979, CFI = .994, NNFI = .990, MFI = .99, RMSEA = .040). These results suggest a good model fit. Additionally, high reliability values were identified for Form B of the instrument, reinforcing its internal consistency (α = .81, ω = .91). These indices indicate strong internal consistency in the measurements obtained for Form B, suggesting substantial reliability within the CFA framework (see Figure 2)

Figure 2.
Measurement Model for Form B: Confirmatory Factor Analysis



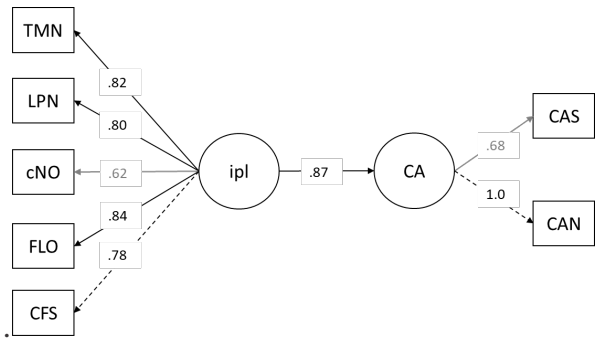
Note.IPL = IPAL-C (July); TMJ = Cloze Task (July); LPJ = Pseudoword Reading (July); CNO = Functional Knowledge of Written Language (July); FLO = Oral Fluency (July); CFS = Phonological Awareness (July); CA = Alphabetic Knowledge (July); CAS = Letter Sound Knowledge (July); CAN = Letter Name Knowledge (July). Source: Own elaboration.

Additional Note: The confirmatory factor analysis (CFA) for Form B demonstrates an adequate fit between the model and the data, with a χ^2 value of 17.909 (p = .161), indicating that the model does not show significant discrepancies with the observed data. Goodness-of-fit indices, such as χ^2/df = 1.377, RMSEA = .040, and CFI = .994, suggest a good model fit, meaning that

the relationships among factors are well represented in the data. Additionally, the indices NFI = .979 and NNFI = .990 further support the model's validity. With respect to reliability, the coefficients α = .81 and ω = .91 indicate strong internal consistency in the measurements, highlighting the notable reliability of Form B of the instrument within the confirmatory factor analysis framework.

In Form C, a χ^2 value of 17.671 (p = .170) was obtained, suggesting an acceptable fit between the model and the data. Considering the goodness-of-fit indices, the model demonstrates a good fit to the empirical data (χ^2/df = 1.359, SRMR = .017, NFI = .984, CFI = .996, NNFI = .993, MFI = .91, RMSEA = .03). These results indicate that the model fits the dataset well. Additionally, consistent reliability values were observed for Form C of the instrument, supporting its internal consistency (α = .83, ω = .92). These indices indicate strong internal consistency in the measurements for Form C, suggesting notable reliability within the CFA framework (see Figure 3)

Figure 3.
Measurement Model for Form C: Confirmatory Factor Analysis



Note. IPL = IPAL-C (November); TMN = Cloze Task (November); LPN = Pseudoword Reading (November); CNO = Functional Knowledge of Written Language (November); FLO = Oral Fluency (November); CFS = Phonological Awareness (November); CA = Alphabetic Knowledge (November); CAS = Letter Sound Knowledge (November); CAN = Letter Name Knowledge (November). Source: Own elaboration.

Additional Note: The confirmatory factor analysis (CFA) for Form C demonstrates an adequate model fit to the data, with a χ^2 value of 17.671 (p = .170), suggesting that the model does not exhibit significant discrepancies with the observed data. Goodness-of-fit indices, such as χ^2/df = 1.359, RMSEA = .03, and CFI = .996, indicate an overall good model fit. The NFI = .984 and NNFI = .993 values further support the model's validity, reflecting its ability to adequately represent the relationships among the factors. With respect to reliability, the coefficients α = .83 and ω = .92 suggest strong internal consistency, reinforcing

the notable reliability of Form C of the instrument within the confirmatory factor analysis framework.

Diagnostic accuracy

To establish the classification of risk status and identify first-grade students at risk, the 25th percentile was used as the threshold. Students who obtained scores ≤ the 25th percentile were classified as at risk, while those with scores > the 25th percentile were considered not at risk. Using this criterion as a reference, the scores obtained by students on the ABCDETI criterion test were dichotomized and recoded as risk = 1, no risk = 0. To analyze the diagnostic accuracy of the IPAL-C, the ROC curve statistical technique was applied, and the following indices were calculated: area under the curve (AUC), sensitivity (Se), specificity (Sp), predictive value (PV), and likelihood ratio (LR). In Table 1, the results related to the classification of risk status based on the ABCDETI test are presented for each of the parallel forms of the IPAL-C, considering the omnibus score for each measurement point (i.e., ZIPALABR_P, ZIPALJUL_P, and ZIPALNOV_P). This omnibus measure was calculated via the arithmetic mean of the standardized scores for each of the reading tasks in the first grade for each parallel form of the IPAL-C (= ZCNL + ZCSL + ZCFS + ZCLE + ZLP + ZTM + ZFLO/7). The mean standardized scores indicate that students classified as at risk scored lower on the IPAL-C measures than those classified as not at risk. In Form A (ZIPALABR_P), 26.1% of the total sample was classified as at risk, equivalent to 68 participants, whereas 73.9% were classified as not at risk, corresponding to 192 participants. In Form B (ZIPALJUL_P), 28% of the total sample was classified as at risk, equivalent to 73 participants, whereas 72% were classified as not at risk, corresponding to 187 participants. In Form C (ZIPALNOV_P), 28.8% of the total sample was classified as at risk, equivalent to 75 participants, whereas 71.2% were classified as not at risk, corresponding to 185 participants.

Table 1.
Classification of Risk Status Based on the Omnibus Measures of the IPAL-C

	Group	Mean	SD	N
ZIPALABR_P	Nonrisk	.23	.61	192
	Risk	-.55	.58	68
	Total	.00	.70	260
ZIPALJUL_P	Nonrisk	.32	.50	187
	Risk	-.76	.62	73
	Total	-.00	.74	260
ZIPALNOV_P	Nonrisk	.37	.49	185
	Risk	-.87	.68	75
	Total	.00	.80	260

Source: Own elaboration.

Table 2 presents the results obtained for all CBMs of the IPAL-C associated with the different cutoff points, according to the criteria of the Youden Index (J), used to calculate sensitivity (Se) and specificity (Sp). The results demonstrated the effectiveness of the IPAL-C measures in identifying both at-risk and non-at-risk students, showing an AUC above .70 at all measurement points. The only exception was the phonological awareness segmentation measure in Form A (AUC_Apr = .44). Measures related to decoding automatization (i.e., CNL, LP, and FLO) and sentence comprehension (i.e., TM) exhibited high classification accuracy. Specifically, LP and FLO achieved an AUC above .90 across all measurement points. Meanwhile, CNL and TM showed an AUC above .80 in Form A, increasing to .90 in Forms B and C. Notably, the cloze task in Form C reached an AUC_Nov = .97. The lowest sensitivity (Se) was observed for the phonological awareness segmentation measure in Form A (CFS = .53), whereas the highest values were found for letter sound knowledge in Form A (CSL_Apr = .93) and the IPAL Z-score in November (zIPAL_NOV_P = .93). Regarding specificity (Sp), all measures exceeded .60, with the highest value observed for the cloze task in Form C (TM_Nov = .94) (see Figures 4-6).

Table 2.
Area under the curve and predictive values associated with the IPAL-C cutoff points

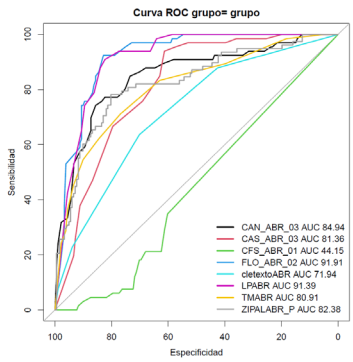
Form	Measure	IPAL-C ABC	Pc	Predictive Values		
				Se	Sp	Criterion
A	CNL	.84	29	.89	.61	4
	CSL	.81	7	.93	.61	4
	CFS	.44	8	.53	.88	
	CLE	.71	4	.63	.70	5
	LP	.91	7	.90	.82	1 y 2
	TM	.80	6	.71	.76	1 y 3
	FLO	.91	20	.92	.82	1 y 2
	zIPAL_P	.82	-.28	.78	.80	4
B	CNL	.90	47	.91	.70	1 y 2
	CSL	.79	15	.79	.67	4
	CFS	.81	18	.80	.66	4
	CLE	.77	3	.61	.81	5
	LP	.90	17	.82	.77	1 y 3
	TM	.94	9	.88	.94	1 y 3
	FLO	.96	39	.92	.80	1 y 2
	zIPAL_P	.90	-.20	.88	.86	1 y 3
C	CNL	.92	68	.88	.84	1 y 3
	CSL	.78	27	.78	.65	4
	CFS	.89	33	.85	.76	1 y 3
	CLE	.81	4	.81	.72	1 y 3
	LP	.90	25	.84	.74	1 y 3
	TM	.97	15	.92	.89	1 y 2
	FLO	.93	60	.91	.79	1 y 2

Form	Measure	IPAL-C ABC	Pc	Predictive Values Se	Sp	Criterion
	zIPAL_P	.92	-.06	.93	.82	1 y 2

Note. CSL = Letter Sound Knowledge; CNL = Letter Name Knowledge; CFS = Phonological Awareness: Segmentation; CLE = Functional Knowledge of Written Language; zIPAL = Global Z score of the IPAL-C in first grade; AUC = Area under the curve; PC = Cutoff points; Se = Sensitivity; Sp = Specificity.

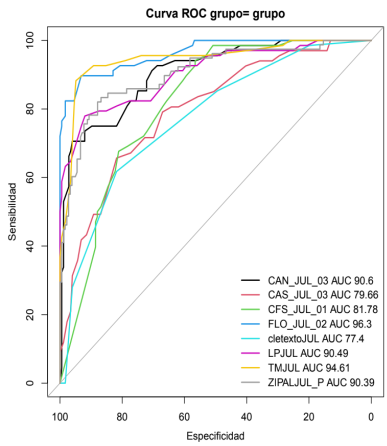
Source: Own elaboration.

Figure 4.
Comparative Graph of the ROC Curves for the IPAL-C in April



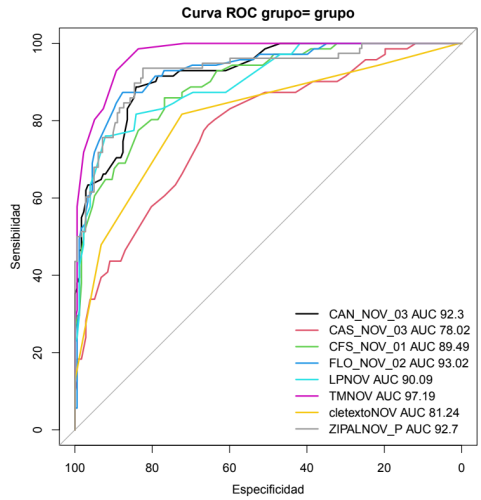
Note. CAN = Letter name knowledge; CAS = Letter sound knowledge; CFS = Phonological awareness; FLO = Oral fluency; LP = Pseudoword reading; TM = Cloze task; CLETEXTO = Functional knowledge of written language; ZIPAL = Omnibus score.
Source: Own elaboration.

Figure 5.
Comparative graph of the ROC curves for the IPAL-C in July



Note. CAN = Letter Name Knowledge; CAS = Letter Sound Knowledge; CFS = Phonological Awareness; FLO = Oral Fluency; LP = Pseudoword Reading; TM = Cloze Task; CLETEXTO = Functional Knowledge of Written Language; ZIPAL = Omnibus Score.
Source: Own elaboration.

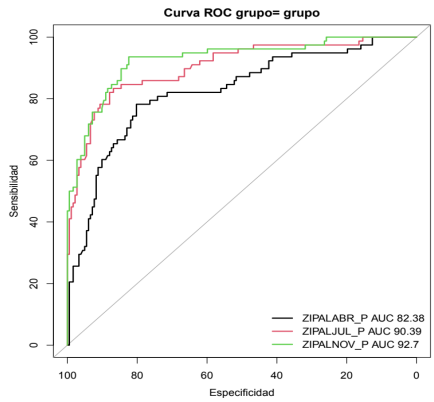
Figure 6.
Comparative graph of the ROC curves for the IPAL-C in November



Note: CAN = Letter name knowledge; CAS = Letter sound knowledge; CFS = Phonological awareness; FLO = Oral fluency; LP = Pseudoword reading; TM = Cloze task; CLETEXTO = Functional knowledge of written language; ZIPAL = Omnibus score.
Source: Own elaboration

The ROC curve, in terms of its classification level for each measure (Z score omnibus measure) that constitutes the IPAL-C in its three forms (ZIPALABR_P, ZIPALJUL_P, and ZIPALNOV_P), shows that Form A presented an AUC greater than 0.80, whereas Forms B and C presented an AUC greater than 0.90, demonstrating a high classification level across all three forms (see Figure 7).

Figure 7.
Comparative Graph of the ROC Curves for the IPAL-C across the Three Measurement Points for the Omnibus Score



Note: ZIPALABR = Omnibus Score April; ZIPALJUL = Omnibus Score July; ZIPALNOV = Omnibus Score November.
Source: Own elaboration.

In terms of predictive values, the IPAL-C demonstrates high classification accuracy, both for students identified as at risk and for those classified as not at risk. The positive predictive values (PPVs) are above 0.40, and in the Omnibus measures, they exceed 0.60. Moreover, the negative predictive values (NPVs) are above 0.80, with two omnibus measures exceeding 0.90. A PPV above 0.40 is considered moderately good, indicating that when the model classifies a student as "at risk," there is a reasonable probability that the student is indeed at risk. PPV values greater than 0.60 in the omnibus measures indicate that the tool has a strong ability to minimize false positives. In other words, students classified as "at risk" have a higher probability of actually being at risk, reducing the number of false positives. An NPV above 0.80 is considered excellent, meaning that most students classified as "not at risk" are truly not at risk, significantly reducing the occurrence of false negatives. The NPV values above 0.90 in the two omnibus measures are exceptionally strong, suggesting that these forms of the IPAL-C are highly effective in accurately identifying students who are not at risk and preventing students who are truly at risk from being incorrectly classified as "not at risk." Table 3 provides a detailed description of the predictive values from the ROC curve. Examining the likelihood ratio (LR), we observe that for the positive likelihood ratio (PLR), all values are above 1, whereas for the negative likelihood ratio (NLR), all values are below 1. This implies that a student with SRD has a high probability of being correctly identified as at risk, whereas the likelihood of a student being classified as at risk is lower if they do not have SRD.

Table 3.
Description of Predictive Values from the ROC Curve

Form	Measure	Predictive Values					
		PPV	PNV	PLR	NLR	FP	FN
	CNL	.46	.94	2.3	.17	69	7
	CSL	.46	.96	2.4	.09	70	4
	CFS	.66	.81	4.6	.52	21	36
A	CLE	.43	.84	2.1	.51	54	24
	LP	.65	.96	5.1	.11	32	6
	TM	.52	.87	3.0	.37	42	19
	FLO	.66	.96	5.3	.09	31	5
	zIPAL_P	.62	.89	3.9	.27	36	17
	CNL	.54	.95	3.0	.12	52	6
	CSL	.47	.89	2.4	.31	58	14
B	CFS	.48	.90	2.4	.28	58	13
	CLE	.56	.84	3.3	.46	32	26
	LP	.58	.91	3.6	.22	39	12

Form	Measure	Predictive Values					
		PPV	PNV	PLR	NLR	FP	FN
	TM	.86	.95	17	.12	9	8
	FLO	.64	.96	4.7	.09	34	5
	zIPAL_P	.73	.92	6.3	.19	24	13
	CNL	.69	.94	5.6	.13	28	8
	CSL	.47	.88	2.2	.32	61	15
C	CFS	.59	.93	3.7	.18	41	10
	CLE	.54	.90	2.9	.25	49	13
	LP	.57	.92	3.3	.20	45	11
	TM	.77	.96	8.6	.07	19	5
	FLO	.63	.95	4.4	.10	37	6
	zIPAL_P	.69	.96	5.3	.07	32	5

Note: PPV = positive predictive value; NPV = negative predictive value; PLR = positive likelihood ratio; NLR = negative likelihood ratio; FP = false positives; FN = false negatives.
Source: Own elaboration.

Discussion and Conclusions

A preventive education approach requires an assessment method capable of accurately measuring students' performance. In this context, the IPAL-C has demonstrated its validity and diagnostic accuracy, both in individual measures and in global measures across the three school-year periods (beginning, middle, and end). This instrument has proven effective in consistently assessing student performance throughout the year, supporting its reliability as a valuable tool in preventive education.

The results of the confirmatory factor analysis (CFA) support a strong model fit, confirming the empirical validity of the data. The latent factor 'reading' explains observable indicators, including alphabetic knowledge, phonological awareness, functional knowledge of written language, pseudoword reading, reading fluency, and comprehension. These findings support the theoretical structure proposed, where the 'reading' factor encompasses and relates to these different components, providing a comprehensive understanding of the reading process. Unlike previous studies using the IPAL in the Canary Islands (Jiménez & García, 2025), where two distinct factors were identified (foundational skills and reading skills), a unidimensional model was found in this sample of Colombian students. This difference could be attributed to variations in early reading instruction, diversity in the educational context, and the methodology used for analysis. In this regard, the skills assessed by the IPAL appear to manifest as a

single integrated continuum in this sample, highlighting the importance of considering the educational context when validating assessment tools.

The results related to risk status classification effectively categorized the students into at-risk and not-at-risk groups. This distinction was reflected in the lower scores obtained by at-risk students in the IPAL-C measure than in those not at risk. This finding highlights the importance of the IPAL-C as a CBM, emphasizing its reliability in effectively identifying students at risk of academic difficulties (Deno, 2003). These results underscore the predictive utility of the IPAL-C in the early identification of potential academic challenges, allowing timely and targeted intervention. The findings revealed that the IPAL-C has a high classification capacity, particularly in the middle and at the end of the school year, with an area under the curve (AUC) above 0.90. When the sensitivity and specificity of the global Z score of the IPAL-C across its parallel forms were evaluated, the sensitivity (identification of at-risk students) exceeded 0.70, and the specificity (identification of not-at-risk students) was above 0.80. Some authors suggest that a reading screening assessment should have a true negative rate (specificity) of at least 80%, and for true positives (sensitivity), they recommend a classification rate above 90% (Compton et al., 2006; Compton et al., 2010). However, according to the National Center on Response to Intervention (2013), such percentages tend to be exceptional for educational tests of this nature.

On the basis of the above, it can be concluded that the validated test in this study demonstrates a "good model fit", as its results reflect strong classification performance and satisfactory capacity to identify both at-risk students and those not at risk, providing a reliable and precise assessment in the educational context.

Additionally, the predictive values of the IPAL-C revealed high accuracy in classifying both at-risk students (true positives - PPV) and not-at-risk students (true negatives - NPV). However, lower accuracy was observed in the classification of false negatives (FNs), indicating that students who were not identified as at risk despite actually being at risk. In terms of prevalence, the IPAL-C demonstrated notable accuracy in identifying students who may have SRD. This was evident through the positive likelihood ratio (PLR) and the negative likelihood ratio (NLR). These results indicate the IPAL-C's ability to identify both true positives and true negatives, although lower accuracy was noted in detecting false negatives, highlighting a potential area for improvement in the detection of at-risk students.

The data suggest that although some at-risk students may perform well on the IPAL-C, their number is relatively small. This classification ensures that the majority of students who could benefit from school support through early intervention actually receive it, whereas only a small percentage do not require additional support. This underscores the importance of using multiple assessment

points, as students not initially classified as at risk in the first evaluation may be identified as at risk in later assessments throughout the school year. This approach implies conducting a continuous student progress monitoring process (Wayman et al., 2012). Among the measures evaluated, pseudoword reading (LP), the cloze task (TM), and oral reading fluency (FLO) stood out for their high classification capacity. However, their effectiveness varied depending on the time of year at which the evaluation was conducted. The only measure that maintained significant predictive power in identifying risk status across all assessment points was oral reading fluency (FLO), which aligns with previous studies in Spanish-speaking samples (Gutiérrez et al., 2021). This highlights FLO as a key indicator in identifying academic difficulties, whereas other measures showed variations in their effectiveness depending on the time of administration during the school year.

In addition to oral reading fluency, this study also emphasized the relevance of pseudoword reading. These findings underscore the importance of phonological processing and reading fluency in assessing reading performance in first grade. In transparent orthographies, such as Spanish, words are relatively easy to read accurately in first grade (Baker et al., 2010; Ellis, 2002). This implies that addressing decoding and fluency properly in the early school years can be a crucial factor in achieving reading comprehension. The strength of CBM assessments such as the IPAL-C lies in their differences from diagnostic assessments on the basis of learning outcomes. CBM tools are designed to be used by school staff without the need for specialists (Elliot et al., 2007; Glover & Albers, 2007). In other words, the IPAL-C provides a tool that can be administered by schools or educational authorities to monitor student learning progress and identify those at risk for SRD. In addition to its ease of administration, the CBM is fast, as each task takes approximately one minute and does not require specialized training. This makes it a useful, easy-to-use, and low-cost tool.

Finally, an important limitation of this study is that the sample was limited to the province of Barranquilla, which limits the generalizability of the findings to other regions of Colombia, given the country's cultural and socioeconomic diversity. Additionally, the sample may not have fully captured the heterogeneity of the population in terms of age, gender, education level, and socioeconomic status. These factors could influence the applicability of the tool in different contexts. Future studies should replicate this research in other Colombian cities, considering various sociodemographic characteristics, to validate the tool in a broader context.

References

Abusamra, V., Chimenti, M. de los A., Difalcis, M., & Vinacur, T. (2024). Lectura y escritura en los primeros años de la escuela primaria: Una sistematización de experiencias de evaluación de la alfabetización inicial en el contexto iberoamericano. Banco Interamericano de Desarrollo (BID).

Acha, J., Rodriguez, N., & Perea, M. (2023). The role of letter knowledge acquisition ability on children's decoding and word identification: Evidence from an artificial orthography. *Journal of Research in Reading, 46*(4), 358–375. https://doi.org/10.1111/1467-9817.12432

Alcantud, F., Alonso, Y., & Rico, D. (2015). Herramientas de cribado para la detección de retrasos o trastornos en el desarrollo: Una revisión sistemática de la literatura. *Revista Española de Discapacidad (REDIS),3*(2), 7-26. http://dx.doi.org/10.5569/2340-5104.03.02.01

Altman, D.G., & Bland, J.M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal, 308*(6943),1552. http://dx.doi.org/10.1136/bmj.308.6943.1552

Anderson, S., Jungjohann, J., & Gebhardt, M. (2020). Effects of using curriculum-based measurement (CBM) for progress monitoring in reading and an additive reading instruction in second classes. *ZFG, 13*, 151 – 161. https://doi.org/10.1007/s42278-019-00072-5

Baker, D.L., Park, Y., & Baker, S.K. (2010). Effect of initial status and growth in pseudoword reading on Spanish reading comprehension at the end of first grade. *Psicothema, 22*(4), 955–962. https://www.ncbi.nlm.nih.gov/pubmed/21044538

Barnett, D. W., Lentz, F. E., Jr., & Macmann, G. (2000). Psychometric qualities of professional practice. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavioral assessment in schools: Theory, research, and clinical foundations* (2nd ed., pp. 355–386). The Guilford Press.

Bravo-Grau, S., & Cruz, J. (2015) Estudios de exactitud diagnóstica: Herramientas para su Interpretación. *Rev. chil. Radiol, 21*(4), 158 -164. http://dx.doi.org/10.4067/S0717-93082015000400007

Cadavid-Ruiz, N., Quijano-Martínez, M., Escobar, P., Rosas, R., & Tenorio, M. (2016). Validación de una prueba computarizada de lectura inicial en niños escolares colombianos. *Ocnos: Revista de Estudios sobre Lectura, 15*(2), 2016, 98-109. https://doi.org/10.18239/ocnos_2016.15.2.1121

Compton, D.L., Fuchs, D., Fuchs, L.S., Bouton, B., Gilbert, J.K., Barquero, L., …& Crouch, R.C. (2010). Selecting at-risk first-grade readers for early intervention:

Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102*(2), 327–340. http://dx.doi.org/10.1037/a0018448

Compton, D.L., Fuchs, D., Fuchs, L.S., & Bryant, J.D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*(2), 394–409. https://doi.org/10.1037/0022-0663.98.2.394

Cortez-Soto, S. N., & Moreno-Treviño, J. O. (2024). Aprendizaje y educación temprana en México: un estudio de descomposición de factores 2012–2018. *Revista Colombiana de Educación, 90*, 185-217. https://doi.org/10.17227/rce.num90-14685

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

Donis, J.H. (2012). Evaluación de la validez y confiabilidad de una prueba diagnóstica. *Avances en Biomedicina, 1*(2),73-81. https://www.redalyc.org/articulo.oa?id=3313/331328015005

Durán-Chiappe, S. M., & Cruz-Velásquez, E. L. (2024). Sentido de la educación inicial: una pesquisa desde los documentos curriculares oficiales. *Revista Colombiana de Educación, 90*, 146-163. https://doi.org/10.17227/rce.num90-14592

Elliot, S.N., Huai, N., & Roach, A.T. (2007). Universal and early screening for educational difficulties: Current and future approaches. *Journal of School Psychology, 45*(2), 137–161. http://dx.doi.org/10.1016/j.jsp.2006.11.002

Ellis, N. (2002). Frequency effects in language processing. A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition, 24* (2), 143 – 188. https://doi.org/10.1017/S0272263102002024

Francis, D.J., Fletcher, J.M., Stuebing, K.K., Lyon, G.R, Shaywitz, B.A., & Shaywitz, S.E. (2005). Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38*(2), 98–108. http://dx.doi.org/10.1177/00222194050380020101

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

Fuchs, L.S., & Fuchs, D. (2011). *Using CBM for progress monitoring in reading*. National Center on Student Progress Monitoring. https://eric.ed.gov/?id=ED519252

Glover, T.A., & Albers, C.A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45*(2), 117–135. https://psycnet.apa.org/doi/10.1016/j.jsp.2006.05.005

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, 6*(33), 1036-1060. https://doi.org/10.1080/10409289.2021.1935537

Hill, T., & Lewicki, P. (2007). STATISTICS: *Methods and Applications*. Stat Soft, Tulsa, OK.

Hintze, J.M., Ryan, A.L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the dynamic indicators of basic early literacy skills and the comprehensive test ofphonological processing. *School Psychology Review, 32*(4), 541–556. https://doi.org/10.1080/02796015.2003.12086220

Jenkins, J.R., & Johnson, E. (2008). *Universal screening for reading problems: Why and how should we do this?* RTI action network.

Jiménez, J.E. (2019). *Modelo de respuesta a la intervención. Un enfoque preventivo para el abordaje de las dificultades específicas de aprendizaje*. Ediciones Pirámide.

Jiménez, J. E., & García Miranda, E. (2025). Evaluación de la validez factorial y la invarianza longitudinal de los indicadores de habilidades básicas de lectura temprana en alumnos españoles de primer grado [Assessing the factorial validity and longitudinal invariance of indicators of basic early reading skills in Spanish first graders]. *Anales de Psicología, 41*(2), 158–171. https://doi.org/10.6018/analesps.615011

Jiménez, J.E., & Gutiérrez, N. (2019). Indicadores de progreso de aprendizaje en lectura [material complementario]. En J.E. Jiménez (Coord.), *Modelo de respuesta a la intervención. Un enfoque preventivo para el abordaje de las dificultades específicas de aprendizaje*. Ediciones Pirámide.

Johnson, E.S., Jenkins, J.R., Petscher, Y., & Catts, H.W. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research and Practice, 24*(4), 174—185. https://doi.org/10.1111/j.1540-5826.2009.00291.x

Johnson, S., & Marlow, N. (2006). Developmental screen or developmental testing? *Early Human Development, 82*(3), 173-183. https://doi.org/10.1016/j.earlhumdev.2006.01.008

Levitt, J.M., Saka, N., Romanelli, L.H., & Hoagwood, K. (2007). Identificación temprana de problemas de salud mental en las escuelas: el estado de la instrumentación. *Psicología Escolar, 45*(2), 163–191. https://doi.org/10.1016/j.jsp.2006.11.005

Liebers, K., Schmidt, C., Junger, R., Prengel, A. (2019). Formatives Assessment in der inklusiven Grundschule im Spannungsfeld von Wissenschaft und Transfer. In: Donie, C., et al. Grundschulpädagogik zwischen Wissenschaft und Transfer. Jahrbuch Grundschulforschung. Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-26231-0_40

McLane, K. (2008). What is curriculum-based measurement and what does it Mean to my child? *National Center on Student Progress Monitoring*. http://www.studentprogress.org

Mellard, D.F., & Johnson, E. (2008). *RTI: A practitioner's guide to implementing response to intervention*. Corwin Press.

National Center on Response to Intervention (2013). Screening Briefs Series—Brief #1: Classification Accuracy. *National Center on Response to Intervention, 2317a* (01), 1- 5. U.S. Department of Education, Office of Special Education Programs. https://eric.ed.gov/?id=ED594240

National Early Literacy Panel. (2009). *Developing early literacy: Report of the National Early Literacy Panel*. National Institute for Literacy. https://lincs.ed.gov/publications/pdf/NELPReport09.pdf

National Reading Panel (US), National Institute of Child Health, & Human Development (US). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and HumanDevelopment, National Institutes of Health. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

Nese, J.F., Anderson, D., Hoelscher, K., Tindal, G., & Alonzo, J. (2011). *Progress Monitoring Instrument Development: Silent Reading Fluency, Vocabulary, and Reading Comprehension. Technical Report # 1110*. Behavioral Research and Teaching, University of Oregon. https://brtprojects.org/wp-

O'Connor, R.E., & Jenkins, J.R. (1999). Prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading, 3*(2), 159–197. https://doi.org/10.1207/s1532799xssr0302_4

Pina-Paredes, V., Harnandez-Pérez, E., Rabadán-Rubio, J.A., Hernández-Pallarés, L., & Fenollar-Cortés, J. (2020). *PREDISCAL Screening de dificultades lectoras y matemáticas*. TEA Ediciones.

Rodríguez, C., Gutiérrez, N., Seoane, R. C., González, D., & de León, S. C. (2024). Reading Spanish prosody: The role of word reading and syntactic knowledge. *Journal of Literacy Research*. *56*(2), 102-122. https://doi.org/10.1177/1086296X241244705

Salvia, J., Ysseldyke, J., & Bolt, S. (2012). *Assessment: In Special and Inclusive Education* (12th Ed). Wadsworth Publishing Co Inc.

Sudaryono., Rahardja, U., Aini, Q., & Isma, Y. (2019). Validity of Test Instruments. *Journal of Physics: Conference Series,* 1–11. https://doi:10.1088/1742-6596/1364/1/012050

UK National Screening Committee. (2009). *Policy review process*. *Programme appraisal criteria*. *Criteria for appraising the viability, effectiveness and appropriateness of a screening programme*. UK National Screening Committee, UK Screening Portal.http://www.screening.nhs.uk/criteria

UK National Screening Committee (2023). *Population screening*. UK National Screening Committee, UK Screening Portal. https://www.gov.uk/guidance/population-screening-explained

Wayman, J.C., Cho, V., Jimerson, J.B., & Spikes, D.D. (2012). District-wide effects on data use in the classroom. *Education Policy Analysis Archives, 20* (25), 1 - 27. https://doi.org/10.14507/epaa.v20n25.2012

Wember, F.B., & Greisbach, M. (2018). Zwischen individueller Leseförderung und inklusivem Unterricht - Didaktische Aspekte der Inklusion im Förderschwerpunkt Lernen. In R. Benkmann & U. Heimlich (Hrsg.), *Inklusion im Förderschwerpunkt Lernen* (S. 133–213). Stuttgart: Kohlhammer.