



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

Recibido: 04 de diciembre de 2023
Evaluado: 03 de enero de 2025
Publicado: 01 de octubre de 2026

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

Tipología: artículo de investigación

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 ABCDE_{ETI} (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

* 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

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

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 ABCD_{ETI} 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

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 ABCD_{ETI} 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

Para citar este artículo:

Villadiego Ramirez, Y., Jiménez, J. E. y Moreno-Torres, M. (2026). Adaptación y validación del IPAL para el tamizaje de habilidades lectoras en estudiantes colombianos de primer grado. *Revista Colombiana de Educación*, (101), e20380, <https://doi.org/10.17227/rce.num101-20380>

Introducción

Las Medidas Basadas en el Currículo (MBC) son herramientas de evaluación diseñadas para medir habilidades fundamentales alineadas con los objetivos de aprendizaje establecidos en el currículo escolar (Fuchs y Fuchs, 2011; MacLane, 2008). Su aplicación en el tamizaje educativo ha demostrado ser eficaz en la detección temprana de dificultades, lo que permite intervenciones oportunas. En este contexto, el presente estudio se centra en la adaptación y validación del IPAL-C, una MBC originalmente desarrollada en España (Jiménez y Gutiérrez, 2019), para su aplicación en estudiantes de primer grado en Colombia. En contraste con las medidas psicométricas tradicionalmente utilizadas en educación, algunas pruebas podrían no ser capaces de demostrar la efectividad del tratamiento o la instrucción (Barnett *et al.*, 2000), y podrían no ser sensibles a los cambios en el proceso de aprendizaje (Salvia *et al.*, 2012). Esto subraya la necesidad de herramientas de evaluación formativa psicométricamente sólidas para la identificación temprana y la intervención. En el ámbito educativo, las pruebas de tamizaje permiten identificar a estudiantes en riesgo de dificultades académicas de manera rápida y eficiente; sin embargo, en el contexto colombiano, existen limitaciones en la disponibilidad de herramientas estandarizadas alineadas con los estándares educativos o lineamientos de aprendizaje. El IPAL-C es una herramienta basada en evidencia, adaptada al español de Colombia y alineada con las necesidades educativas locales, cuya implementación facilita la detección temprana de dificultades lectoras en primer grado. El uso de herramientas de evaluación alineadas con el currículo escolar es fundamental para garantizar la equidad en el acceso a la educación y la identificación temprana de dificultades en el aprendizaje. En este sentido, estudios previos han resaltado la importancia de contar con instrumentos estandarizados que permitan evaluar el desempeño académico desde las primeras etapas educativas (Durán-Chiappe y Cruz-Velásquez, 2024). Asimismo, investigaciones sobre la educación temprana han mostrado que las brechas en el aprendizaje pueden mitigarse a través de estrategias de evaluación y tamizaje que identifiquen de manera oportuna a los estudiantes en riesgo y faciliten intervenciones eficaces (Cortez-Soto y Moreno-Treviño, 2024). Estos hallazgos refuerzan la relevancia del IPAL-C como una herramienta que no solo permite monitorear el progreso lector de los estudiantes, sino que también contribuye a la formulación de políticas educativas basadas en evidencia. Estas pruebas asisten a los educadores en la evaluación del nivel de conocimientos y habilidades de los estudiantes, lo que faculta la identificación de posibles brechas en su aprendizaje. Asimismo, posibilitan el beneficio de tratamientos o evaluaciones adicionales para mitigar estas brechas (UK National Screening Committee, NSC,

2009). Tal como se señala en el informe de Abusamra *et al.* (2024), la implementación de herramientas de evaluación fundamentadas teóricamente y con normas adecuadas es esencial para detectar tempranamente las dificultades en la alfabetización inicial, lo que a su vez activa acciones específicas para su mitigación en el ámbito iberoamericano.

El IPAL ha sido validado previamente en España, donde ha demostrado su utilidad como MBC para el tamizaje de dificultades lectoras y su sensibilidad al cambio en el aprendizaje (Gutiérrez *et al.*, 2021). Dada esta base empírica, el presente estudio tiene como objetivo adaptar y validar el IPAL-C en el contexto colombiano, así como evaluar su estructura factorial y precisión diagnóstica en estudiantes de primer grado. A través de esta validación, buscamos determinar su capacidad para identificar de manera confiable a estudiantes en riesgo de dificultades en la lectura y su alineación con los estándares o lineamientos de aprendizaje del Ministerio de Educación Nacional (MEN) de Colombia. De manera similar, en el ámbito de la prevención de las dificultades de aprendizaje (DA), se ha sugerido que, cuanto antes se identifiquen e intervengan los factores que preceden la manifestación de un problema o trastorno conductual, o de rendimiento académico, mayor será la probabilidad de manejarlos de manera efectiva o incluso mitigarlos. Esto no solo previene que dichos problemas afecten gravemente el rendimiento académico, sino que también contribuye al bienestar emocional de los estudiantes (Levitt *et al.*, 2007; Pina-Paredes *et al.*, 2020). En este sentido, la investigación sobre las MBC representa una base valiosa, teórica y empírica, para la creación de pruebas de tamizaje en lectura, escritura y matemáticas (Jiménez, 2019), las cuales se basan en los objetivos y contenidos específicos del currículo establecido por las autoridades educativas. Las MBC se administran en diferentes momentos del año escolar (inicio-mitad-final) para evaluar el progreso de los estudiantes, identificar áreas en las que necesiten apoyo adicional o incluso informar sobre la eficacia de las intervenciones educativas implementadas. En tanto que se proporcionan datos objetivos y medibles sobre el rendimiento de los estudiantes, estas pruebas permiten a los educadores tomar decisiones informadas sobre la enseñanza y el aprendizaje, al obtener una visión general del nivel de competencia de los estudiantes y detectar de manera temprana a aquellos que puedan necesitar intervención educativa adicional (Anderson *et al.*, 2020; Fuchs *et al.*, 2004; Nese *et al.*, 2011). En el ámbito del aprendizaje de la lectura, el National Reading Panel (NRP) (2000) resaltó la importancia de identificar de manera temprana las dificultades específicas en el aprendizaje de la lectura (DEAL) y los problemas de alfabetización. Sin un apoyo intensivo de lectura, las bajas habilidades de lectura pueden causar importantes problemas de aprendizaje en la escuela secundaria (Wember y Greisbach, 2018).

Evaluación basada en el currículo y lectura

En el contexto de la investigación sobre la validez de las Medidas Basadas en el Currículo en Lectura (MBC-L), se evidencia una extensa literatura científica. En el ámbito de los países de habla hispanoparlante, se han llevado a cabo investigaciones, particularmente en España (Gutiérrez *et al.*, 2021). Debido a la complejidad del proceso lector, las investigaciones han demostrado que el uso de medidas que evalúan un solo componente de la lectura no es beneficioso en la detección de riesgo; por el contrario, una combinación de medidas que evalúen diferentes componentes puede mejorar la precisión del tamizaje (Compton *et al.*, 2006; O'Connor y Jenkins, 1999). Además, las mediciones deben ser sensibles al desarrollo de la lectura, los componentes de la lectura y el nivel de grado en el que se pretende aplicar la prueba; por ejemplo, en primer grado, las medidas de detección más exitosas combinan los siguientes componentes: fluidez para nombrar letras e identificar el sonido, segmentación de fonemas, vocabulario y fluidez en la identificación de palabras (Acha *et al.*, 2023; Compton *et al.*, 2006; Francis *et al.*, 2005; Jenkins y Johnson, 2008; O'Connor y Jenkins, 1999). De hecho, estudios recientes han demostrado que los niños leen con expresión solo cuando han alcanzado un alto nivel de automatización en la lectura (Rodríguez *et al.*, 2024).

En este sentido, es importante tener en cuenta, durante el diseño e implementación de una prueba de tamizaje, la necesidad de someterla a un proceso explícito, sistemático y transparente, el cual implica evaluar la eficacia del tamizaje y determinar su factibilidad (UK National Screening Committee, 2023). Una prueba de tamizaje debe cumplir con múltiples criterios de calidad para ser considerada satisfactoria (Johnson y Marlow, 2006), algunos de los cuales incluyen: a) costo-efectividad, lo que implica la simplicidad, facilidad y brevedad en la administración de la prueba, su bajo costo de uso, accesibilidad, así como su adaptabilidad a diversas lenguas y culturas (Alcantud *et al.*, 2015); en resumen, la prueba debe ser asequible y eficiente en términos de tiempo y recursos requeridos para su administración y la evaluación de sus resultados. En el caso de la lectura, estas pruebas han demostrado ser muy eficaces (Fuchs *et al.*, 2004). b) Fiabilidad: los resultados de la prueba deben ser consistentes, lo que significa que al realizar la prueba en las mismas condiciones y con los mismos sujetos, debería generar resultados similares cada vez, con lo cual se aseguran resultados estables. Para evaluar la confiabilidad de las MBC, se utilizan pruebas como el test-retest, la fiabilidad interjueces y la consistencia interna. En el caso del tamizaje conformado por medidas paralelas aplicadas al inicio, mitad y final del año escolar, se debe establecer la comparabilidad entre las formas. Esto se evalúa a través del coeficiente de fiabilidad de formas paralelas; además, se considera la desviación media y estándar de cada forma (Glover y Albers, 2007). c) Validez: la prueba debe medir de

manera precisa y confiable la variable que se pretende evaluar, respaldada por una sólida evidencia de que mide lo que se supone que debe medir.

Respecto a la validez de las pruebas de tamizaje, es crucial considerar tres tipos de validez: la de constructo, la de contenido y la de criterio (Sudaryono *et al.*, 2019). La primera nos permite saber si un instrumento realmente mide el constructo teórico o la variable que queremos medir, para lo cual, es necesario analizar la estructura interna y externa del instrumento. La validez de contenido indica si los ítems que componen un instrumento de tamizaje reflejan la naturaleza multidimensional de lo que se pretende medir; y la validez de criterio es una medida que nos dice si una evaluación predice qué tan bien se desempeñará la persona evaluada en un criterio particular. A su vez, existen dos tipos de validez de criterio: concurrente y predictiva (Glover y Albers, 2007). La validez concurrente nos permite evaluar hasta qué punto una medida es capaz de identificar a las personas que actualmente tienen dificultades mediante una prueba estandarizada como referencia o prueba criterio (Johnson *et al.*, 2009); mientras que la validez predictiva faculta la distinción entre aquellos que no están en riesgo y aquellos que probablemente tengan dificultades en rendimiento académico o en el comportamiento, según lo que se desea evaluar. De acuerdo con las características de las pruebas de tamizaje, la validez predictiva se considera el criterio más importante a cumplir (Glover y Albers, 2007). Los índices de sensibilidad y especificidad, valores predictivos y los índices de verosimilitud se utilizan a menudo para evaluar la utilidad predictiva del tamizaje (Bravo-Grau y Cruz, 2015).

Un sistema de educación basado en la prevención requiere de un tamizaje que permita medir con precisión a los estudiantes en riesgo de acuerdo con su rendimiento actual, en el área de aprendizaje para la que fue diseñada la MBC (esto es, matemáticas, lectura, escritura, o ciencias) y al grado escolar. Por lo tanto, la precisión diagnóstica es esencial para las pruebas de tamizaje en educación, por las siguientes razones: a) identificación temprana de problemas: una alta precisión en las pruebas de tamizaje permite detectar de manera temprana posibles dificultades o necesidades de los estudiantes, lo que facilita la intervención oportuna para abordar esos problemas. Se considera que ninguna prueba clasifica al 100 % de los sujetos en riesgo (sensibilidad) y no riesgo (especificidad); por esta razón, lo que se busca es poder reducir a un número mínimo los casos que son mal clasificados (Hill y Lewicki, 2007). b) Intervención personalizada: con resultados precisos, los educadores pueden monitorear el progreso del estudiante, adaptar las estrategias de enseñanza y proporcionar apoyo específico a los estudiantes que lo necesitan, lo que mejora sus oportunidades de aprendizaje y desarrollo (Deno, 2003; Fuchs y Fuchs, 2011; Liebers *et al.*, 2019). c) Uso eficiente de recursos: las pruebas de tamizaje precisas ayudan a asignar recursos de manera eficiente al identificar a los estudiantes que requieren atención adicional, con lo cual se evita la inversión inútil en aquellos que no la necesitan. Hintze *et al.* (2003) consideran que, en el contexto del tamizaje, los falsos positivos son generalmente menos problemáticos; lo importante es detectar los que

se encuentran en riesgo, por eso es tan importante un alto nivel de sensibilidad en la prueba. Y d) evaluación del impacto de intervenciones: la precisión diagnóstica también permite medir con mayor precisión el efecto de las intervenciones educativas, lo que es esencial para mejorar continuamente los programas y políticas educativas (Mellard y Johnson , 2008).

Metodología

Objetivos e hipótesis

Este estudio se centra en examinar la validez de constructo del IPAL-C y determinar si tanto el IPAL-C en su totalidad como sus componentes relacionados con la lectura son capaces de identificar a estudiantes en riesgo y no riesgo en el primer grado de primaria. Las hipótesis planteadas incluyen: a) la adaptación del IPAL al contexto colombiano (IPAL-C) presentará una sólida validez de constructo de forma similar al encontrado en estudios previos con muestras españolas; b) las medidas paralelas del IPAL-C demostrarán una precisión diagnóstica adecuada para clasificar a estudiantes en riesgo y no riesgo de presentar DEAL; c) el conjunto de medidas del IPAL-C mostrará una mayor capacidad de precisión diagnóstica en comparación con las medidas individuales por separado.

Participantes

La muestra para este estudio estuvo conformada por 292 estudiantes de primer grado de básica primaria (155 niños, 137 niñas; edad en meses, $M = 79,8$; $DT = 6,66$), provenientes de cinco escuelas del distrito de Barranquilla, Colombia. Barranquilla es la capital del departamento del Atlántico, en la región Caribe del país, y se divide en cinco localidades: Suroccidente, Suroriente, Norte-Centro Histórico, Metropolitana y Riomar. Para garantizar la representatividad geográfica, se seleccionó un colegio por cada una de estas localidades, y se conformó una muestra diversa en términos de estratos socioeconómicos y características regionales. La muestra estuvo compuesta por tres escuelas públicas y dos privadas. En cuanto a su clasificación, según el desempeño académico, de acuerdo con el Instituto Colombiano para la Evaluación de la Educación (Icfes), los colegios en Colombia se agrupan en cinco categorías según los resultados de las pruebas Saber 11: A+, A, B, C y D, siendo A+ las de más alto desempeño y D las de menor rendimiento. En el presente estudio, una escuela correspondió a la categoría A+ (muy superior), dos a la categoría A (superior), una a la categoría B (bajo) y una a la categoría D (inferior). Se tomaron los grupos completos de clase para su participación, para un total de diez grupos

evaluados, con dos grupos por cada escuela. Aquellos estudiantes que presentaban un diagnóstico de trastornos del neurodesarrollo o discapacidades (esto es, sensorial, física, motora o intelectual) fueron excluidos del estudio. Las escuelas seleccionadas representan una diversidad socioeconómica y cultural. En consecuencia, el nivel educativo de los padres varía, lo que refleja la heterogeneidad de antecedentes educativos y recursos en Barranquilla.

Instrumentos

La adaptación de los Indicadores de Progreso de Aprendizaje en Lectura (IPAL) (Jiménez y Gutiérrez, 2019) al contexto colombiano abarcó no solo aspectos idiomáticos, sino también la alineación entre el IPAL y los estándares educativos o lineamientos de aprendizaje de primer grado propuesto por el MEN en cuanto al proceso lector. Esta adaptación, al igual que su versión original, cuenta con tres formas paralelas de evaluación, a saber, Forma A (inicio), Forma B (mitad) y Forma C (final), aplicadas en tres momentos diferentes del año escolar (abril, julio y noviembre). El IPAL cuenta con un manual de instrucciones, cuadernillo del estudiante y plantilla del examinador, y se hace necesario contar con un cronómetro para su aplicación. Esta MBC-L mide las habilidades que el National Reading Panel (2000) identifica como principales componentes de lectura: conocimiento alfabético, vocabulario, conciencia fonémica, comprensión y fluidez. A estos cinco componentes se suma el componente de conocimientos funcionales del lenguaje escrito, que, según el informe del National Early Literacy Panel (2009), ha demostrado ser uno de los mejores predictores de la comprensión lectora en la educación inicial. Las tareas incluidas en el IPAL son las siguientes: a) Fluidez en el Conocimiento del Nombre de las Letras (CNL): los niños nombran tantas letras en mayúsculas y minúsculas como puedan en un minuto. Se puntúa por el número de letras identificadas correctamente. b) Fluidez en el Conocimiento del Sonido de las Letras (CSL): los estudiantes identifican letras y sus sonidos en un minuto. Se puntúa por el número de sonidos identificados correctamente. c) Conciencia Fonológica de Segmentar (CFS): los estudiantes segmentan palabras sin sentido en fonemas en un minuto, con dificultad progresiva. Se puntúa por el número de fonemas identificados correctamente. d) Conocimientos funcionales acerca del Lenguaje Escrito (CLE): evaluación del conocimiento básico del lenguaje impreso con seis preguntas sobre la estructura y uso de libros de texto. Se puntúa por el número de respuestas correctas. e) Lectura de Pseudopalabras (LP): los estudiantes leen palabras sin sentido en un minuto para evaluar su habilidad en la correspondencia grafema-fonema. Se puntúa por el número de palabras leídas correctamente. f) Textos Mutilados (TM): los estudiantes completan frases incompletas seleccionando la opción correcta entre tres alternativas en cinco minutos. Se puntúa por el número de respuestas

correctas. g) Fluidez en Lectura Oral (FLO): los estudiantes leen un texto breve en voz alta durante un minuto, con lo cual se miden velocidad y precisión. Se puntúa por el número de palabras leídas correctamente.

El test ABCDeti (Cadavid-Ruiz et al., 2016) se utilizó como prueba-criterio para identificar el estatus de riesgo de presentar o no DEAL en los estudiantes de primer grado. Esta prueba adaptada y validada en Colombia para los grados jardín, primero y segundo, evalúa, a través de un programa computarizado, ocho habilidades que los autores consideran centrales en el proceso lector y que están recogidas en el informe National Reading Panel (2000). Su aplicación se hizo junto con la prueba de final del año escolar del IPAL-C (Forma C), en noviembre. El test ABCDeti evalúa las siguientes habilidades en los estudiantes de primer grado de básica primaria: principio alfabético, conciencia fonológica, tipos de textos, fluidez y precisión lectora, comprensión lectora y comprensión oral.

Procedimiento

Se revisó la herramienta IPAL desarrollada para escolares en las Islas Canarias (España), adaptada al contexto colombiano (IPAL-C), mediante un proceso que incluyó: a) ajuste de vocabulario y estructuras lingüísticas para garantizar su adecuación a la modalidad del español de Colombia, b) alineación con los estándares educativos o lineamientos de aprendizaje de lenguaje para primer grado definido por el MEN de Colombia, y c) evaluación de la equivalencia conceptual y psicométrica de los ítems en relación con la versión original. El presente estudio adoptó un diseño longitudinal en el que el IPAL-C fue administrado en tres momentos durante el año académico con el propósito de monitorear el progreso en habilidades lectoras y evaluar su precisión diagnóstica en la identificación de estudiantes en riesgo. La Forma A se aplicó después de esperar un tiempo de adaptación y progreso escolar, aproximadamente un mes y medio posterior al inicio del año escolar; las Formas B y C del IPAL-C se aplicaron con tres meses de diferencia entre una y otra. En cuanto a la prueba criterio ABCDeti, esta se aplicó junto con la aplicación de la Forma C del IPAL-C en el mes de noviembre, con el objetivo de obtener la validez concurrente. Las respuestas de cada estudiante para cada medida fueron registradas en la hoja de evaluación del examinador, para las tres formas del IPAL-C y la prueba ABCDeti. Todas las pruebas fueron administradas de manera estandarizada por evaluadores previamente capacitados en el protocolo de aplicación del IPAL-C.

Análisis de datos

Para evaluar la estructura interna del IPAL-C y su validez de constructo, se utilizó un Análisis Factorial Confirmatorio (AFC) en cada uno de los tres momentos de

medición, para cada uno de los cuales se especificaron como parámetros libres las varianzas de los términos de error y se fijó a uno un coeficiente estructural en la variable latente o factor. El método de estimación empleado en cada análisis fue el de Máxima Verosimilitud Robusta (MLR), y se calcularon los índices de fiabilidad a través de los coeficientes alfa de Cronbach (α) y omega jerárquico (Ω). Para evaluar el ajuste de los modelos, se emplearon como medidas absolutas de ajuste el estadístico Chi-cuadrado y la Raíz Cuadrática Media (RMSEA); como medidas de ajuste incremental, el Índice Tucker-Lewis (TLI), y el índice de bondad de ajuste comparativo (CFI) para comparar las diferencias entre los residuales de la matriz de covarianza de la muestra y el modelo hipotetizado. Este análisis permitió determinar si las tareas que componen el IPAL-C reflejan adecuadamente un modelo teórico unidimensional de la competencia lectora, en el que las habilidades evaluadas están asociadas a un único factor latente de lectura.

Para determinar la precisión diagnóstica del IPAL-C en la identificación de estudiantes en riesgo, se empleó el análisis de la Curva Receiver Operating Characteristic (ROC) y se calculó el Área Bajo la Curva (ABC) y los índices de sensibilidad y especificidad. Este método permitió establecer la capacidad del instrumento para clasificar correctamente a los estudiantes en riesgo de dificultades lectoras. La curva ROC es una representación gráfica que evalúa la capacidad de un instrumento para distinguir entre grupos (por ejemplo, individuos con y sin riesgo). La curva muestra la relación entre la sensibilidad (verdaderos positivos) y 1-especificidad (falsos positivos) en diferentes puntos de corte. El área bajo la curva (ABC) mide la precisión global del instrumento, donde un valor de 1 indica una discriminación perfecta y un valor de 0,5 sugiere que no es mejor que el azar. Así, se establece en un principio un percentil específico tomando como referencia la prueba criterio administrada junto con la prueba de tamizaje (Johnson et al., 2009). De acuerdo con lo anterior, para la clasificación del estatus de riesgo, se calculó la puntuación global, tanto de la prueba criterio empleada ABCDeti, como del IPAL-C. La puntuación global del ABCDeti se convirtió en una variable dicotómica, riesgo y no riesgo, con lo cual se estableció el percentil 25 para la clasificación de los estudiantes en situación de riesgo. El ABC nos permite probar el rendimiento del modelo de clasificación de datos para ver si se puede distinguir con precisión entre los estudiantes que tienen riesgo de presentar DEAL y los que no. Los criterios se basaron en el índice de Youden (J) para seleccionar el punto de corte que mejor clasificaba a los estudiantes en riesgo. El índice J es una métrica utilizada en el análisis de sensibilidad y especificidad para identificar el punto de corte óptimo de un instrumento. Este índice se calcula como $J = \text{Sensibilidad} + \text{Especificidad} - 1$ = $\text{Sensibilidad} + \text{Especificidad} - 1$, y su valor máximo indica el punto que optimiza simultáneamente la sensibilidad y especificidad, con lo cual se equilibran ambos aspectos. Gracias a estas mediciones podemos determinar la calidad y precisión del tamizaje en términos de sensibilidad y especificidad.

La sensibilidad es el porcentaje de estudiantes identificados con precisión como en riesgo, mientras que la especificidad es el porcentaje de estudiantes identificados con precisión como en no riesgo. Los resultados de las pruebas de tamizaje colocan a los estudiantes en uno de cuatro grupos: Verdadero Positivo (VP), Verdadero Negativo (VN), Falso Positivo (FP) y Falso Negativo (FN) (National Center on Response to Intervention, 2013). Una vez determinada la sensibilidad y la especificidad, se calcularon además los siguientes índices: Valor Predictivo (VP) y Razón de Verosimilitud (RV). Se hace necesario estimar también la probabilidad en que la identificación realizada por el instrumento sea correcta, por lo que se evalúa el Valor Predictivo Positivo (VPP) y el Valor Predictivo Negativo (VPN) (Donis, 2012). La razón de verosimilitud, o eficiencia pronóstica, también permite determinar la precisión de la prueba de tamizaje, pero indica la posibilidad (prevalencia) del desarrollo de la afección; es decir, permite saber si los resultados de las pruebas de tamizaje aumentan o disminuyen la probabilidad de estar en riesgo de presentar DEAL. Hay dos tipos de razón de verosimilitud: Razón de Verosimilitud Positiva (RVP) y Razón de Verosimilitud Negativa (RVN) (Altman y Bland, 1994; Donis, 2012).

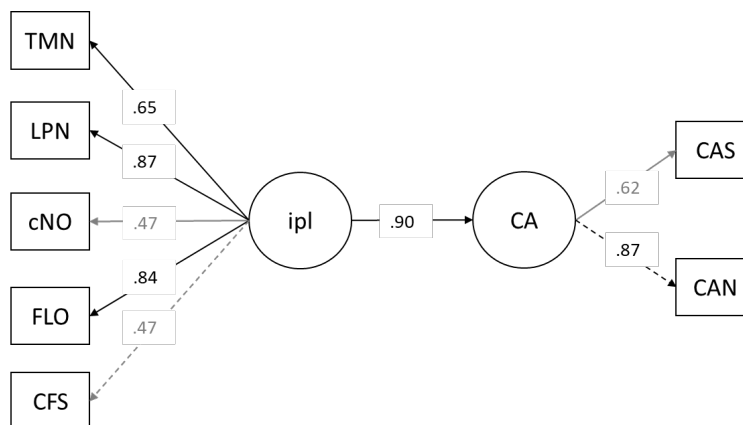
Resultados

Se analizó la estructura factorial de cada forma paralela de primer grado, a través del AFC. A partir de este análisis, se propone un modelo hipotético de la competencia lectora, basado en los datos empíricos, obtenidos por medio de las respuestas que los estudiantes dieron al emplear como instrumento la herramienta IPAL-C. Bajo este modelo y a la luz de la revisión teórica, se sostiene la presencia de un único factor de primer orden relacionado con la competencia lectora. Por lo tanto, todas las tareas incluidas en el IPAL-C (como CNL, CSL, CFS, CLE, LP, TM y FLO) en los tres momentos del año escolar (abril, julio y noviembre) están vinculadas a ese factor latente. Este modelo refleja una concepción teórica unidimensional de la competencia lectora.

En la forma A, se obtiene el valor de χ^2 de 17,234 ($p = .189$). Esto sugiere que el modelo y los resultados se ajustan entre sí. Al analizar los índices de bondad de ajuste obtenidos durante la estimación del modelo, se observa que este se ajusta bien a los datos ($SRMR = 0,031$; $\chi^2=17,234$; $df = 13$; $p = .189$; $NFI = .975$; $CFI = .994$; $NNFI = .990$; $MFI = .991$; $RMSEA = .036$). Los resultados indican una buena fiabilidad de la Forma A del IPAL-C en el modelo factorial confirmatorio, lo que demuestra una consistencia interna adecuada, $\alpha = .71$, $\omega = .85$. Estos valores sugieren una consistencia interna aceptable en la medición hecha para la escala evaluada en el contexto del modelo factorial confirmatorio (figura 1).

Figura 1.

Modelo de medida forma A. Análisis factorial confirmatorio.



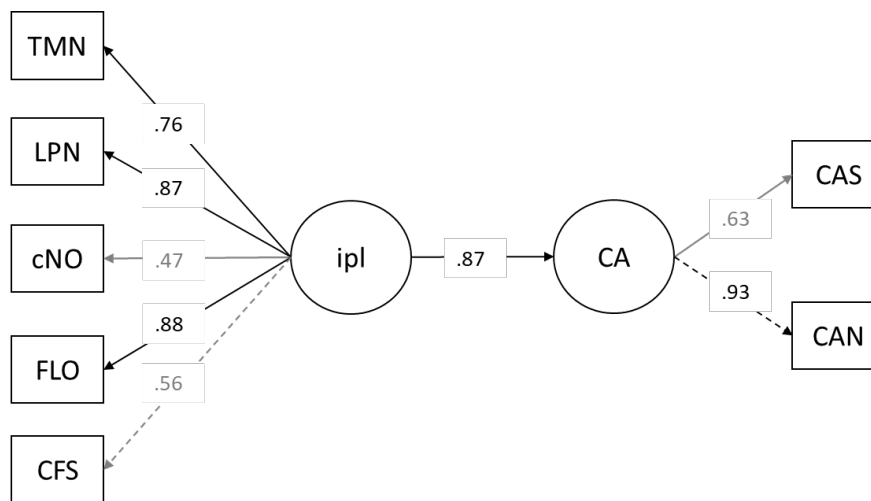
Nota: IPL= IPAL-C (abril); TMA = Textos Mutilados (abril); LPA = Lectura de Pseudopalabras (abril); cNO = Conocimientos Funcionales Lenguaje Escrito (abril); FLO = Fluidez Oral (abril); CFS = Conciencia Fonológica (abril); CA = Conocimiento Alfabético (abril); CAS: Conocimiento del Sonido de las Letras (abril); CAN = Conocimiento del Nombre de las Letras (abril).
Fuente: elaboración propia.

*Nota adicional: el análisis factorial confirmatorio para la Forma A muestra un ajuste adecuado entre el modelo y los datos, respaldado por un χ^2 de 17,234 ($p = .189$), que indica que las discrepancias entre el modelo y los datos no son estadísticamente significativas. Los índices de bondad de ajuste como el SRMR = .031, el CFI = .994 y el RMSEA = .036 están dentro de los valores recomendados, lo que refuerza la validez del modelo. Además, el índice NFI = .975 y el NNFI = .990 confirman que el modelo es apropiado para los datos. La fiabilidad de la Forma A también es adecuada, con coeficientes de $\alpha = .71$ y $\omega = .85$, lo que sugiere una consistencia interna satisfactoria en las mediciones realizadas. Estos resultados indican que el modelo de la Forma A tiene un buen ajuste y una fiabilidad adecuada, lo que valida su uso en el análisis factorial confirmatorio.

En la Forma B, se obtuvo un valor de χ^2 de 17,909 ($p = .161$), lo que indica un ajuste razonable entre el modelo y los datos. Al evaluar los índices de bondad de ajuste, se observa que el modelo se ajusta adecuadamente a los datos empíricos ($\chi^2/df = 1,377$, SRMR = .025, NFI = .979, CFI = .994, NNFI = .990, MFI = .99, RMSEA = .040). Estos resultados sugieren un buen ajuste del modelo. Además, se identificaron valores de fiabilidad altos para la Forma B del instrumento, lo que fortalece la consistencia interna de este, $\alpha = .81$, $\omega = .91$. Estos índices indican una consistencia interna destacada en las mediciones realizadas para la Forma B del instrumento, lo cual sugiere una fiabilidad sustancial en el contexto del AFC (figura 2).

Figura 2.

Modelo de medida forma B. Análisis factorial confirmatorio.



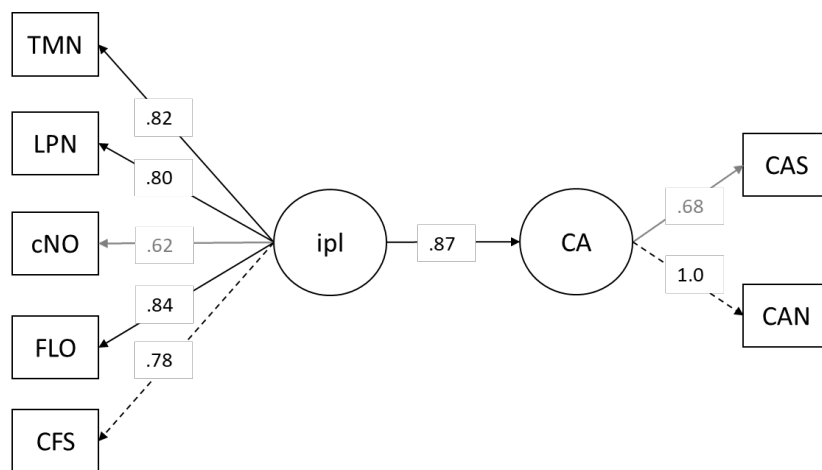
Nota: IPL= IPAL-C (julio); TMJ = Textos Mutilados (julio); LPJ = Lectura de Pseudopalabras (julio); cNO = Conocimientos Funcionales Lenguaje Escrito (julio); FLO = Fluidez Oral (julio); CFS = Conciencia Fonológica (julio); CA = Conocimiento Alfabético (julio); CAS: Conocimiento del Sonido de las Letras (julio); CAN = Conocimiento del Nombre de las Letras (julio).
Fuente: elaboración propia.

*Nota adicional: el análisis factorial confirmatorio para la Forma B muestra un ajuste adecuado entre el modelo y los datos, con un χ^2 de 17,909 ($p = .161$), lo que indica que el modelo no presenta discrepancias significativas con los datos observados. Los índices de bondad de ajuste como el $\chi^2/df = 1,377$, el RMSEA = .040 y el CFI = .994 sugieren un buen ajuste del modelo, lo que significa que las relaciones entre los factores se representan adecuadamente en los datos. Adicionalmente, los índices NFI = .979 y NNFI = .990 refuerzan la validez del modelo. En cuanto a la fiabilidad, los coeficientes $\alpha = .81$ y $\omega = .91$ indican una consistencia interna robusta en las mediciones, lo que resalta la fiabilidad destacada de la Forma B del instrumento en el contexto del análisis factorial confirmatorio.

En la Forma C, se obtuvo un valor de χ^2 de 17,671 ($p = .170$), lo que sugiere un ajuste aceptable entre el modelo y los datos. Al considerar los índices de bondad de ajuste, se evidencia que el modelo se ajusta bien a los datos empíricos ($\chi^2/df = 1,359$, SRMR = .017, NFI = .984, CFI = .996, NNFI = .993, MFI = .91, RMSEA = .03). Estos resultados sugieren un buen ajuste del modelo al conjunto de datos. Además, se observaron valores consistentes de fiabilidad para la Forma C del instrumento, lo que respalda la consistencia interna de esta versión, $\alpha = .83$, $\omega = .92$. Estos índices indican una sólida consistencia interna en las mediciones realizadas para la Forma C del instrumento, lo que sugiere una fiabilidad destacada en el contexto del AFC (figura 3).

Figura 3.

Modelo de medida forma C. Análisis factorial confirmatorio.



Nota: IPL= IPAL-C (noviembre); TMN = Textos Mutilados (noviembre); LPN = Lectura de Pseudopalabras (noviembre); cNO =Conocimientos Funcionales Lenguaje Escrito (noviembre); FLO = Fluidez Oral (noviembre); CFS = Conciencia Fonológica (noviembre); CA = Conocimiento Alfabético (noviembre); CAS: Conocimiento del Sonido de las Letras (noviembre); CAN = Conocimiento del Nombre de las Letras (noviembre).

Fuente: elaboración propia.

*Nota adicional: el análisis factorial confirmatorio de la Forma C muestra un ajuste adecuado del modelo a los datos, con un χ^2 de 17,671 ($p = .170$), lo que sugiere que el modelo no presenta discrepancias significativas con los datos observados. Los índices de bondad de ajuste, como el $\chi^2/df = 1,359$, el RMSEA = .03 y el CFI = .996, indican un buen ajuste general del modelo. Los valores de NFI = .984 y NNFI = .993 también apoyan la validez del modelo, lo que refleja su capacidad para representar de manera adecuada las relaciones entre los factores. En cuanto a la fiabilidad, los coeficientes $\alpha = .83$ y $\omega = .92$ sugieren una consistencia interna sólida, lo que refuerza la fiabilidad destacada de la Forma C del instrumento en el contexto del análisis factorial confirmatorio.

Precisión diagnóstica

Para establecer la clasificación del estatus de riesgo, y detectar el estudiantado de primer grado en riesgo de presentar DEAL, se empleó el percentil 25, frente a lo cual se consideraron en riesgo aquellos estudiantes que obtuvieran puntuaciones $\leq$ percentil 25, y en no riesgo a aquellos con puntuaciones superiores $>$ percentil 25. Al tomarse como referencia este criterio, se llevó a cabo la dicotomización de las puntuaciones obtenidas por los estudiantes en la prueba criterio ABCDeti, siendo recodificadas como riesgo = 1, no riesgo = 0.

Para analizar la capacidad o precisión diagnóstica del IPAL-C, se empleó la técnica estadística Curva ROC y se calcularon los índices: Área Bajo la Curva (ABC), sensibilidad (Se), especificidad (Ep), Valor Predictivo (vp) y Razón De Verosimilitud (RV). En la tabla 1 se presentan los resultados vinculados a la clasificación del estatus de riesgo a partir del Test ABCDeti, para cada una de las formas paralelas IPAL-C, que consideran la puntuación ómnibus para cada momento de medida (esto es, ZIPALABR_P, ZIPALJUL_P y ZIPALNOV_P). Esta medida ómnibus se calculó a partir de la media aritmética de las puntuaciones típicas de cada una de las tareas para primer grado para cada forma paralela del IPAL-C ($= ZCNL + ZCSL + ZCFS + ZCLE + ZLP + ZTM + ZFLO/7$). Las puntuaciones medias típicas muestran que los estudiantes en situación de riesgo presentan más bajas puntuaciones en las medidas del IPAL-C, que aquellos estudiantes que no están en riesgo. En la forma A (ZIPALABR_P), el 26,1 % de la muestra total fue clasificado en riesgo, lo que equivale a 68 participantes; mientras que el 73,9 % de los evaluados se clasificaron en no riesgo, lo que corresponde a 192 participantes. En la forma B (ZIPALJUL_P), el 28 % de la muestra total fue clasificado en riesgo, lo que equivale a 73 participantes; mientras que el 72 % de los evaluados se clasificaron en no riesgo, lo que corresponde a 187 participantes. En la forma C (ZIPALNOV_P), el 28,8 % de la muestra total fue clasificado en riesgo, lo que equivale a 75 participantes, mientras que el 71,2 % de los evaluados se clasificaron en no riesgo, lo que corresponde a 185 participantes.

Tabla 1.

Clasificación del estatus de riesgo con las medidas Ómnibus del IPAL-C

	Grupo	Media	DT	N
ZIPALABR_P	No-riesgo	.23	.61	192
	Riesgo	-.55	.58	68
	Total	.00	.70	260
ZIPALJUL_P	No-riesgo	.32	.50	187
	Riesgo	-.76	.62	73
	Total	-.00	.74	260
ZIPALNOV_P	No-riesgo	.37	.49	185
	Riesgo	-.87	.68	75
	Total	.00	.80	260

Fuente: elaboración propia.

En la tabla 2 se muestran los resultados obtenidos para todas las MBC del IPAL-C asociadas a los distintos puntos de corte, de acuerdo con los criterios del Índice de Youden (J), para calcular la sensibilidad (Se) y la especificidad (Ep). Los resultados mostraron la tasa de éxito de las medidas del IPAL-C tanto para detectar a los estudiantes en riesgo como a aquellos que no lo están. Se observa un ABC superior a .70 en todos los momentos de medida; solamente se encuentra un valor por debajo, en la forma A, la medida conciencia fonológica de segmentar (ABCCabr= .44). Las medidas relacionadas con la automatización en la decodificación (esto es, CNL, LP y FLO) y la comprensión de frases (esto es, TM) mostraron una alta capacidad de clasificación. Para el caso de LP y FLO, el ABC fue superior a .90 en todos los momentos de medida; por su parte, CNL y TM mostraron una clasificación por encima de .80 en la forma A, y un aumento a .90 en las formas B y C; de hecho, textos mutilados en la forma C muestra un ABCnov = .97. La sensibilidad (Se) más baja la mostró conciencia fonológica de segmentar en la forma A (CFS = .53); la más alta fue conocimiento del sonido de las letras en la forma A (CSLabr= .93) y la puntuación Z del IPAL en noviembre (zIPALNOV_P = .93). En cuanto a la especificidad (Ep), todas las medidas se encuentran por encima de .60; siendo la más alta textos mutilados en la forma C (TMnov= .94) (figuras 4-6).

Tabla 2.

Área Bajo la Curva y valores predictivos asociados a los puntos de corte del IPAL-C

Formas	Medida	IPAL-C	PC	Valores predictivos		
		ABC		Se	Ep	Criterio
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

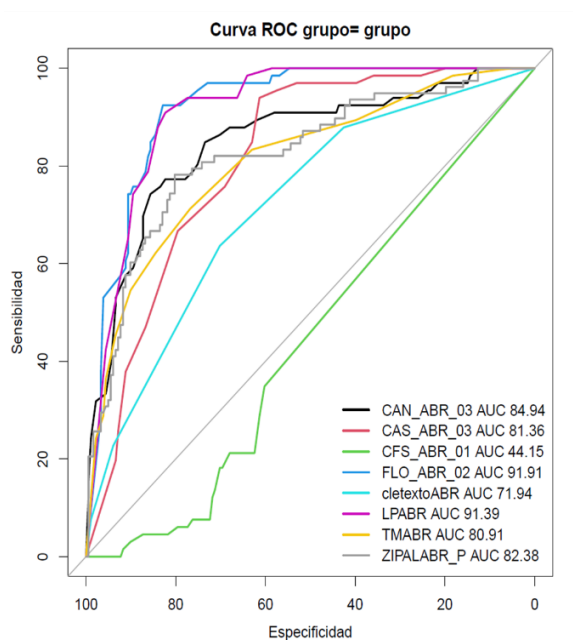
C	CNL	.92	68	.88	.84	1 y 3
	CSL	.78	27	.78	.65	4
	CFS	.89	33	.85	.76	1 y 3
	CLE	.81	4	.81	.72	1 y 3
	LP	.90	25	.84	.74	1 y 3
	TM	.97	15	.92	.89	1 y 2
	FLO	.93	60	.91	.79	1 y 2
	zIPAL_P	.92	-.06	.93	.82	1 y 2

Nota: CSL = Conocimiento del Sonido de las Letras; CNL = Conocimiento del Nombre de las Letras; CFS= Conciencia Fonológica: Aislar; CLE= Conocimientos funcionales acerca de Lenguaje Escrito; zIPAL = puntuación Z global del IPAL-C en primer grado; ABC= Área Bajo la Curva; PC = Puntos de Corte; Se= sensibilidad; Ep= especificidad.

Fuente: elaboración propia.

Figura 4.

Gráfica comparativa de las curvas ROC del IPAL-C en abril.

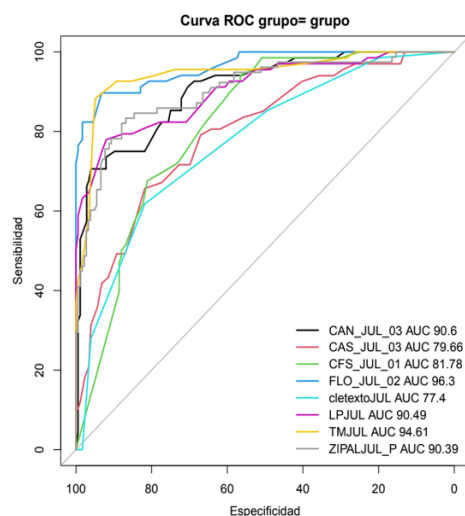


Nota: CAN = Conocimiento del Nombre de las Letras; CAS = Conocimiento del Sonido de las Letras; CFS = Conciencia Fonológica; FLO = Fluidez Oral; LP = Lectura de Pseudopalabras; TM = Textos Mutilados; CLETEXTO = Conocimientos Funcionales del Lenguaje Escrito; ZIPAL = Puntuación Omnibus.

Fuente: elaboración propia.

Figura 5.

Gráfica comparativa de las curvas ROC del IPAL-C en julio.

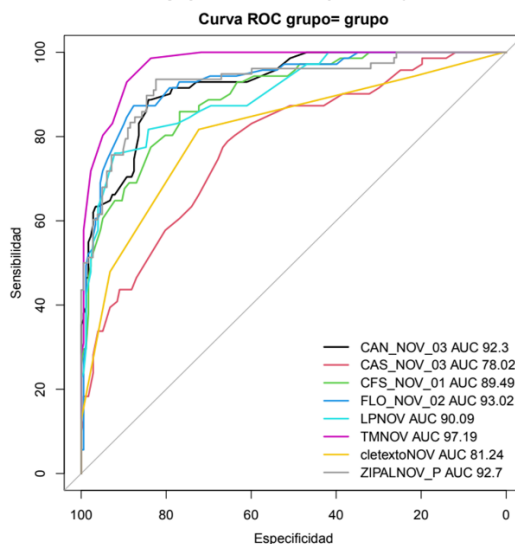


Nota: CAN = Conocimiento del Nombre de las Letras; CAS = Conocimiento del Sonido de las Letras; CFS = Conciencia Fonológica; FLO = Fluidez Oral; LP = Lectura de Pseudopalabras; TM = Textos Mutilados; CLETEXT = Conocimientos Funcionales del Lenguaje Escrito; ZIPAL = Puntuación Ómnibus.

Fuente: elaboración propia.

Figura 6.

Gráfica comparativa de las curvas ROC del IPAL-C en noviembre.



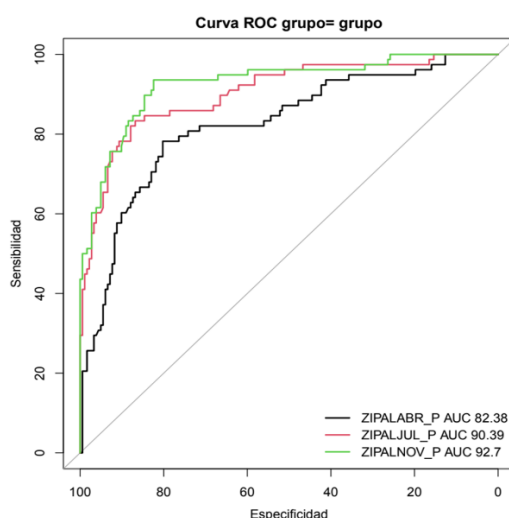
Nota: CAN = Conocimiento del Nombre de las Letras; CAS = Conocimiento del Sonido de las Letras; CFS = Conciencia Fonológica; FLO = Fluidez Oral; LP = Lectura de Pseudopalabras; TM = Textos Mutilados; CLETEXT = Conocimientos Funcionales del Lenguaje Escrito; ZIPAL = Puntuación Ómnibus.

Fuente: elaboración propia.

En cuanto a su nivel de clasificación para cada medida (puntuación Z-Medida Ómnibus) que conforma el IPAL-C en sus tres formas (ZIPALABR_P, ZIPALJUL_P y ZIPALNOV_P), la curva ROC evidencia que la forma A mostró un ABC superior a 0,80, mientras que las formas B y C presentaron un ABC superior a 0,90, lo que demuestra un alto nivel de clasificación en las tres formas (figura 7).

Figura 7.

Gráfica comparativa de las curvas ROC del IPAL-C en los tres momentos de medida para la puntuación ómnibus.



Nota: ZIPALABR = Puntuación Ómnibus Abril; ZIPALJUL = Puntuación Ómnibus Julio; ZIPALNOV = Puntuación Ómnibus Noviembre.

Fuente: elaboración propia.

En cuanto a los valores predictivos, el IPAL-C demuestra un porcentaje alto de clasificación correcta, tanto para estudiantes considerados en riesgo, como de aquellos que no se encuentran un riesgo. Los valores predictivos positivos (VPP) se observan por encima de 0,40, y en las medidas Ómnibus, por encima de .60; asimismo, los valores predictivos negativos (VPN) están por encima de 0,80 y, en dos de las medidas ómnibus, incluso se ubica por encima de 0,90. Un VPP por encima de 0,40 es moderadamente bueno e indica que, cuando el modelo clasifica a un estudiante como “en riesgo”, hay una probabilidad razonable de que efectivamente esté en riesgo. Los VPP superiores a 0,60 en las medidas Ómnibus indican una buena capacidad de la herramienta para evitar falsos positivos; es decir, los estudiantes clasificados como “en riesgo” tienen una mayor probabilidad de estar realmente en riesgo, lo que minimiza el número de falsos positivos. Un VPN superior a 0,80 es

excelente, lo que implica que la mayoría de los estudiantes clasificados como “no en riesgo” realmente no lo están, lo que reduce significativamente los falsos negativos. Los VPN por encima de 0,90 en dos de las medidas Ómnibus son excepcionalmente buenos, lo que sugiere que estas formas del IPAL-C son muy efectivas para identificar correctamente a los estudiantes que no están en riesgo y evitar que aquellos que realmente están en riesgo sean clasificados incorrectamente como “no en riesgo”. La tabla 3 muestra una descripción de los valores predictivos de la curva ROC. Al revisar la razón de verosimilitud (RV), observamos que para la razón de verosimilitud positiva (RVP), todos los valores están por encima de 1, mientras que para la razón de verosimilitud negativa (RVN), todos los valores están por debajo de 1. Esto implica que un estudiante que presenta DEAL tiene una alta probabilidad de ser detectado como en riesgo, mientras que la probabilidad de que un estudiante sea detectado como en riesgo es menor cuando no presenta DEAL.

Tabla 3.

Descripción de valores predictivos de la curva de ROC

Formas	Medida	Valores Predictivos					
		VPP	VPN	RVP	RVN	FP	FN
A	CNL	.46	.94	2.3	.17	69	7
	CSL	.46	.96	2.4	.09	70	4
	CFS	.66	.81	4.6	.52	21	36
	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
B	CNL	.54	.95	3.0	.12	52	6
	CSL	.47	.89	2.4	.31	58	14
	CFS	.48	.90	2.4	.28	58	13
	CLE	.56	.84	3.3	.46	32	26
	LP	.58	.91	3.6	.22	39	12
	TM	.86	.95	17	.12	9	8
	FLO	.64	.96	4.7	.09	34	5
	zIPAL_P	.73	.92	6.3	.19	24	13
C	CNL	.69	.94	5.6	.13	28	8
	CSL	.47	.88	2.2	.32	61	15
	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

Formas	Medida	Valores Predictivos					
		VPP	VPN	RVP	RVN	FP	FN
	FLO	.63	.95	4.4	.10	37	6
	zIPAL_P	.69	.96	5.3	.07	32	5

Nota: VPP = Valor Predictivo Positivo; VPN= Valor Predictivo Negativo; RVP = Razón de Verosimilitud Positiva; RVN = Razón de Verosimilitud Negativa; FP = Falsos Positivos; FN = Falsos Negativos.

Fuente: elaboración propia.

Discusión y conclusiones

Un enfoque educativo centrado en la prevención requiere de un método de evaluación que pueda medir con precisión el rendimiento de los estudiantes. En este contexto, el IPAL-C ha demostrado su validez y precisión diagnóstica, tanto en cada una de las medidas de forma individual, como en las medidas globales a lo largo de los tres momentos del año escolar (inicio, mitad y final). Este instrumento se ha mostrado efectivo para evaluar de manera consistente el rendimiento de los estudiantes a lo largo del año, lo que respalda su utilidad como una herramienta confiable en el contexto educativo preventivo.

Los resultados del Análisis Factorial Confirmatorio (AFC) respaldan un ajuste sólido del modelo propuesto, lo que confirma la validez empírica de los datos. Se ha observado que el factor latente ‘lectura’ explica los indicadores observables, como el conocimiento alfabético, la conciencia fonológica, el conocimiento sobre el lenguaje escrito, la lectura de pseudopalabras, la fluidez lectora y la comprensión. Estos hallazgos respaldan la estructura teórica propuesta, a partir de la cual el factor ‘lectura’ engloba y se relaciona con estos diferentes componentes, lo que proporciona una comprensión integral del fenómeno lector. A diferencia de estudios previos con el IPAL en las Islas Canarias (Jiménez y García, 2025), donde se identificaron dos factores diferenciados (habilidades fundamentales y habilidades lectoras), en esta muestra de estudiantes colombianos se encontró un modelo unidimensional. Esta diferencia podría deberse a variaciones en la instrucción temprana de la lectura, la diversidad en el contexto educativo y la metodología de análisis utilizada. En este sentido, las habilidades evaluadas por el IPAL parecen manifestarse como un continuo integrado en esta muestra, lo que resalta la importancia de considerar el contexto educativo al validar herramientas de evaluación.

Los resultados relacionados con la clasificación del estado de riesgo evidenciaron una categorización eficaz de los estudiantes en grupos de riesgo y sin riesgo. Esta

distinción se reflejó en las puntuaciones más bajas obtenidas por los estudiantes clasificados en riesgo en la medida IPAL-C, en comparación con aquellos que no presentaban riesgo. Este hallazgo destaca la importancia del IPAL-C como MBC y su confiabilidad para identificar de manera efectiva a los estudiantes en riesgo de dificultades académicas (Deno, 2003). Así mismo, subraya la utilidad y la capacidad predictiva del IPAL-C en la detección temprana de posibles desafíos académicos, a fin de una intervención oportuna y específica para estos estudiantes. Los resultados mostraron que el IPAL-C tiene una alta capacidad de clasificación, especialmente en la mitad y al final del curso escolar, con un Área Bajo la Curva (ABC) superior a 0,90. Al evaluar la sensibilidad y especificidad de la puntuación Z global del IPAL-C en sus formas paralelas, se identificó una sensibilidad (estudiantes en riesgo) superior a 0,70 y una especificidad (estudiantes sin riesgo) superior a 0,80. Algunos autores sugieren que una prueba de tamizaje en lectura debería mostrar una tasa de verdaderos negativos (especificidad) mayor o igual al 80 % y, para los verdaderos positivos (sensibilidad), sugieren un porcentaje superior al 90 % en la clasificación (Compton *et al.*, 2006; Compton *et al.*, 2010). Sin embargo, según el National Center on Response to Intervention (2013), estos porcentajes tienden a ser excepcionales en pruebas educativas de este tipo.

Con base en lo anterior, se puede concluir que la prueba validada en este estudio muestra una “buena bondad de ajuste”, ya que sus resultados reflejan un sólido desempeño en la clasificación y una capacidad satisfactoria para identificar tanto a estudiantes en riesgo como a aquellos que no lo están, lo que ofrece una evaluación confiable y precisa en el ámbito educativo.

Además, los valores predictivos del IPAL-C revelaron una alta precisión en la clasificación tanto de estudiantes en riesgo (VPP) como de aquellos que no están en riesgo (VPN). Sin embargo, se observaron valores más bajos en la clasificación de estudiantes categorizados como falsos negativos (FN), a saber, aquellos que no fueron identificados como en riesgo a pesar de estarlo. En cuanto a la prevalencia, el IPAL-C demostró una precisión notable al identificar a los estudiantes que podrían presentar dificultades específicas en el aprendizaje de la lectura (DEAL). Esto se evidenció mediante la razón de verosimilitud positiva (RVP) y la razón de verosimilitud negativa (RVN). Estos resultados indican la capacidad del IPAL-C para identificar tanto a los verdaderos positivos como a los verdaderos negativos, aunque se observó una menor precisión en la identificación de los falsos negativos, lo que señala un área de mejora en la detección de estudiantes en riesgo de DEAL.

Los datos sugieren que, aunque algunos estudiantes en riesgo pueden tener un rendimiento óptimo en el IPAL-C, su número es relativamente reducido. Esta clasificación asegura que la gran mayoría de los estudiantes que podrían beneficiarse del apoyo escolar a través de una intervención temprana realmente lo necesiten, mientras que solo un pequeño porcentaje no lo requerirá. Lo anterior subraya la

importancia de introducir más de una medida, ya que los estudiantes que no son clasificados inicialmente como en riesgo en la primera evaluación podrían ser identificados como en riesgo en evaluaciones posteriores a lo largo del año académico. Dicho enfoque implica llevar a cabo un proceso continuo de monitoreo del progreso del estudiantado (Wayman *et al.*, 2012).

Entre las medidas evaluadas, se observó que la Lectura de Pseudopalabras (LP), los Textos Mutilados (TM) y la Fluidez en Lectura Oral (FLO) destacaron por su alta capacidad de clasificación de los estudiantes; no obstante, su efectividad varió según el momento del año escolar en que se aplicó la evaluación. La única medida que mantuvo una contribución significativa para predecir el estatus de riesgo fue la Fluidez en Lectura Oral (FLO), lo cual coincide con estudios previos en muestras españolas (Gutiérrez *et al.*, 2021). Esto resalta la importancia de esta medida como un indicador clave en la identificación del riesgo de dificultades académicas, mientras que las otras medidas mostraron variaciones en su efectividad según el momento de aplicación durante el año escolar.

Además de este predictor, en este estudio se destaca la relevancia de la lectura de pseudopalabras. Estos resultados subrayan la importancia del procesamiento fonológico y la fluidez lectora para evaluar el desempeño en lectura durante el primer grado. En ortografías transparentes o poco profundas, como el español, las palabras son relativamente fáciles de leer con precisión en el primer grado (Baker *et al.*, 2010; Ellis, 2002), de lo que se deduce que un adecuado abordaje de la decodificación y su fluidez durante los primeros grados escolares en los establecimientos educativos puede ser un factor crucial para lograr la comprensión lectora. La fortaleza de este tipo de pruebas radica en su diferencia con las evaluaciones diagnósticas basadas en el producto del aprendizaje. Las pruebas MBC, como el IPAL-C, están diseñadas para ser utilizadas por el personal escolar y no requieren la presencia de especialistas (Elliot *et al.*, 2007; Glover y Albers, 2007). En otras palabras, el IPAL-C proporciona una herramienta que puede ser manejada por la escuela o las Secretarías de Educación para evaluar el progreso de aprendizaje de los estudiantes e identificar aquellos en riesgo de DEAL. Además de su fácil aplicabilidad, las pruebas son rápidas, ya que constan de tareas que toman aproximadamente un minuto y no exigen ningún tipo de capacitación especializada, lo que las convierte en una herramienta útil, fácil de usar y de bajo costo.

Por último, una limitación importante de este estudio es que la muestra se centró exclusivamente en la provincia de Barranquilla, lo que limita la generalización de los resultados a otras regiones de Colombia, dada la diversidad cultural y socioeconómica del país. Además, la muestra podría no haber reflejado completamente la heterogeneidad de la población en términos de variables como edad, género, nivel educativo y estrato socioeconómico, lo que podría influir en la aplicabilidad de la herramienta en diferentes contextos. Futuros estudios deberían

replicar esta investigación en otras ciudades colombianas, y considerar diversas características sociodemográficas, para validar la herramienta en un contexto más amplio.

Referencias

- Abusamra, V., Chimenti, M. de los A., Difalcis, M. y 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., Rodríguez, N. y 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. 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. y 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. y 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. 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. y Macmann, G. (2000). Psychometric qualities of professional practice. En: E. S. Shapiro y T. R. Kratochwill (Eds.), *Behavioral assessment in schools: Theory, research, and clinical foundations* (segunda edición), (pp. 355-386). The Guilford Press.
- Bravo-Grau, S. y Cruz, J. (2015) Estudios de exactitud diagnóstica: Herramientas para su Interpretación. *Revista Chilena de Radiología*, 21(4), 158-164. <http://dx.doi.org/10.4067/S0717-93082015000400007>

- Cadavid-Ruiz, N., Quijano-Martínez, M., Escobar, P., Rosas, R. y 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. y 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. y 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. y 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. y 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. y 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. y 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. y 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. y 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. y 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. y 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. y Lewicki, P. (2007). *Statistics: Methods and Applications*. Stat Soft, Tulsa, OK.
- Hintze, J. M., Ryan, A. L. y Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the dynamic indicators of basic early literacy skills and the comprehensive test of phonological processing. *School Psychology Review*, 32(4), 541-556. <https://doi.org/10.1080/02796015.2003.12086220>
- Jenkins, J. R. y 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. y 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. 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. y 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. y 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. y Prengel, A. (2019). Formatives Assessment in der inklusiven Grundschule im Spannungsfeld von Wissenschaft und Transfer. En: 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*. www.studentprogress.org
- Mellard, D. F. y 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. 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, y 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 Human Development, National Institutes of Health. <https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/reports/rt.pdf>
- Nese, J. F., Anderson, D., Hoelscher, K., Tindal, G. y 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. y 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., Hernandez-Pérez, E., Rabadán-Rubio, J. A., Hernández-Pallarés, L. y Fenollar-Cortés, J. (2020). *Prediscal Screening de dificultades lectoras y matemáticas*. TEA Ediciones.
- Rodríguez, C., Gutiérrez, N., Seoane, R. C., González, D. y 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. y Bolt, S. (2012). *Assessment: In Special and Inclusive Education* (doceava edición). Wadsworth Publishing Co Inc.
- Sudaryono, Rahardja, U., Aini, Q. y Isma, Y. (2019). Validity of Test Instruments. *Journal of Physics: Conference Series*, 1-11. <https://doi.org/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. y 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., y Greisbach, M. (2018). Zwischen individueller Leseförderung und inklusivem Unterricht-Didaktische Aspekte der Inklusion im Förderschwerpunkt Lernen. En: R. Benkmann y U. Heimlich (Comp.), *Inklusion im Förderschwerpunkt Lernen* (pp. 133-213). Kohlhammer.