

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), enero-febrero 2026,
Volumen 10, Número 1.

https://doi.org/10.37811/cl_rcm.v10i1

**EVIDENCIA BASADA EN EL MODELO LATENTE DE
LA BATERÍA TBA-P: ANÁLISIS FACTORIAL Y
CONSISTENCIA INTERNA EN APRENDIZAJES
INSTRUMENTALES, INTELIGENCIA FACTORIAL Y
FUNCIONES EJECUTIVAS**

**EVIDENCE BASED ON THE LATENT MODEL OF THE TBA-P
BATTERY: FACTOR ANALYSIS AND INTERNAL CONSISTENCY IN
INSTRUMENTAL LEARNING, FACTORIAL INTELLIGENCE AND
EXECUTIVE FUNCTIONS**

Pablo Jesús San Martín Catalán
Centro de Estudios Psicométricos y del Aprendizaje

Evidencia basada en el modelo latente de la Batería TBA-P: Análisis factorial y consistencia interna en aprendizajes instrumentales, inteligencia factorial y funciones ejecutivas

Pablo Jesús San Martín Catalán¹

psanmartin@cepaonline.cl

<https://orcid.org/0000-0002-2230-1635>

Centro de Estudios Psicométricos y del Aprendizaje

RESUMEN

El presente trabajo tiene como objetivo aportar evidencia psicométrica para un modelo de evaluación psicopedagógica de alta precisión, tomando como vector el Test Básico de Aprendizaje Profundo. Primero, se desarrolla un análisis crítico de instrumentos ampliamente utilizados en contextos clínicos y escolares en Chile, examinando sus fundamentos teóricos, procedimientos psicométricos y sistemas de puntuación. Segundo, se discuten las limitaciones de baterías con enfoque curricular y de modelos clásicos que, aun presentando adecuada fiabilidad y validez, utilizan sistemas de puntuación aritméticos simples, lo que genera distorsiones en la representación de los constructos latentes, especialmente en perfiles heterogéneos de relevancia clínica. Tercero, se describe el diseño conceptual y psicométrico del TBA-P como una batería multidimensional e informatizada de medición profunda, estructurada sobre pruebas unidimensionales que discriminan entre procesos cognitivos, aprendizajes primitivos y aprendizajes instrumentales. Se presenta evidencia de fiabilidad mediante el coeficiente KR-20 y evidencia de validez factorial para los entornos cognitivo-ejecutivo y lecto-ortográfico, a través de análisis factorial exploratorio y confirmatorio. Por último, se expone un modelo de puntuación factorial-estructural que permite una estimación más sensible de las variables latentes evaluadas.

Palabras clave: test psicopedagógico, inteligencia general, aprendizajes instrumentales, aprendizajes primitivos, psicometría

¹ Autor principal

Correspondencia: psanmartin@cepaonline.cl

Evidence based on the latent model of the TBA-P Battery: Factor analysis and internal consistency in instrumental learning, factorial intelligence and executive functions

ABSTRACT

This paper aims to provide psychometric evidence for a highly accurate psychoeducational assessment model, using the Basic Test of Deep Learning (TBA-P) as its basis. First, a critical analysis of instruments widely used in clinical and school settings in Chile is presented, examining their theoretical foundations, psychometric procedures, and scoring systems. Second, the limitations of curriculum-focused batteries and classical models are discussed. These models, despite exhibiting adequate reliability and validity, use simple arithmetic scoring systems, which distort the representation of latent constructs, especially in clinically relevant heterogeneous profiles. Third, the conceptual and psychometric design of the TBA-P is described as a multidimensional, computerized battery for in-depth measurement, structured around unidimensional tests that discriminate between cognitive processes, primitive learning, and instrumental learning. Reliability evidence is presented using the KR-20 coefficient, and factorial validity evidence for the cognitive-executive and reading-orthographic domains is presented through exploratory and confirmatory factor analysis. Finally, a structural-factorial scoring model is presented that allows for a more sensitive estimation of the latent variables assessed.

Keywords: psychoeducational test, general intelligence, instrumental learning, primitive learning, psychometrics

*Artículo recibido 02 enero 2026
Aceptado para publicación: 30 enero 2026*



INTRODUCCION

Sobre la necesidad de una batería de evaluación psicométrica-psicopedagógica de alta precisión

Antes de profundizar todo tipo de concepto asociado a la medición del aprendizaje humano y los procesos cognitivos que hacen posible dicho fenómeno, es necesario establecer un contexto definicional acerca de la Psicometría y la Psicopedagogía teórica (San Martín, 2025a), por lo cual, en este trabajo, todo lo expuesto versa en la definición de Psicometría (a), medición (b) y modelo psicopedagógico de medición (c):

- a. “Psicometría puede definirse en términos generales como el conjunto de métodos, técnicas y teorías implicadas en la medición de variables psicológicas” (Muñiz, 2018, p.15). En este sentido, en coherencia con este trabajo, en términos particulares y de profundización epistemológica, puede definirse como disciplina científica dedicada al estudio de la medición de variables cognitivas latentes y no latentes.
- b. De acuerdo con San Martín (2025a), medición es la cuantificación de un fenómeno cognitivo variable.
- c. El modelo psicopedagógico de medición (M) describe la relación entre el test (T) (instrumento usado para medir), el puntaje (X), la variable (V) (latente y no latente), el sujeto (S) (aquel poseedor de variables), los errores de medida (e) y los factores influyentes (F). A razón de esto, “El test (T) no es solo un instrumento de medición, sino el sistema dentro del cual opera M. En otras palabras, M estructura la relación entre S y V, pero es dentro de T donde se implementa este modelo para procesar $V + e$ y obtener X, que representa la estimación empírica del atributo o rasgo medido” (San Martín, 2025a, p.37).

Referir a error de medida es fundamental en el campo de la evaluación psicométrica, ya que este fija los rangos probables en los cuales se encontrará el puntaje más representativo de la cognición y aprendizaje del sujeto evaluado. En esta lógica, el modelo que subyace es el de Spearman, donde todo puntaje obtenido mediante un test es información mental condicionada por el error de medida, lo cual se traduce como -puntaje empírico-. Matemáticamente, la formalización es:

$$X = V + e$$



Si se calculan las varianzas para una definición clara, tenemos que:

$$\sigma_e^2 = \sigma_X^2 - \sigma_V^2$$

Con lo cual, si consideramos que cuando el error es cero la fiabilidad es total (1), mediante deducción tenemos que:

$$f = \frac{\sigma_V^2}{\sigma_V^2 + \sigma_e^2}$$

Por lo tanto, al despejar el error queda:

$$\sigma_e^2 = \sigma_X^2 - \sigma_X^2 f$$

Y si se entiende que la varianza es el cuadrado de la desviación estándar, al factorizar se obtiene el error estándar como:

$$e = \sigma_X \sqrt{1 - f}$$

En consecuencia, como argumento se puede decir que, a mayor fiabilidad, menor error asociado a la medición, por lo cual es totalmente relevante elegir el instrumento que genere el menor error de medida, entendiéndose bajo el supuesto que el test está diseñado para aquello que se pretende medir, constituyendo además validez para la interpretabilidad.

Los instrumentos psicométricos clínicos y escolares precisan de una serie de demostraciones científicas para ratificar la validez y precisión, entendiéndose como validez el grado en que un test mide aquello que dice medir. Por ejemplo, si un instrumento se sustenta en la hipótesis para evaluar el desarrollo lector como unidad y al mismo tiempo los ítems y variables asociadas no se explican a modo de varianza por el desarrollo lector (variable latente o no observable), no se constituiría como instrumento válido para dicho fin. La precisión en cambio, o fiabilidad, es el grado de homogeneidad de los resultados asociados a un sujeto, es decir, cuán consistente internamente es el test (sus ítems) y/o cuán similares son los resultados de distintas aplicaciones al mismo sujeto.

Para estudiar con mayor rigor este asunto, se analizarán tres pruebas psicométricas usadas en Chile en espacios formales como centros clínicos y escolares, usando como referencia los documentos de diseño, estandarización y/o propiedades psicométricas.



Batería psicopedagógica EVALÚA

Uno de los test más usados hasta entonces es la batería psicopedagógica EVALÚA. Este test estandarizado es de origen español y ha sido sujeto de cambios para la adaptación chilena, dichos cambios no han sido justificados en un artículo científico como habitualmente se hace en este contexto. Las versiones van de 1.0 a 4.0, siendo esta última la que se utiliza desde el año 2019 (Vidal & González Manjón, 2019). EVALÚA es un instrumento especialmente escolar, con enfoque curricular (niveles por curso), pero, no ajustada al currículum chileno propiamente tal, o al menos no hay evidencia al respecto. Si se analiza el alcance evaluativo del test EVALÚA, usando como ejemplo el EVALÚA-6 (Vidal & González Manjón, 2019), se obtiene:

- A. En la página 26 del manual, se plantean las correlaciones entre las puntuaciones y el rendimiento escolar de los estudiantes ($r=0,77$), aludiendo a una muestra mayor que 1000, sin embargo, no hay ninguna referencia ni detalle del estudio, no se describe el procedimiento ni los alcances y tampoco se describe la organización muestral.
- B. En la página 26 del manual, respecto a la escala de matemáticas, se plantea que la correlación entre el test de cálculo y numeración y el test de resolución de problemas es de 0.615, lo cual contradice lo que dice el mismo manual en la página 112, cuyo resultado informado es una correlación de Pearson de 0,177, mediante una muestra de más de 60 mil estudiantes. Esto es problemático dado que no queda clara la relación respecto a las variables que componen una escala y en consecuencia una dimensión de análisis.
- C. En la página 26-27 se observa el resultado de un análisis factorial exploratorio (AFE) sin mostrar el cumplimiento de los requisitos matemáticos para este tipo de análisis. No se calcula la adecuación muestral (KMO) ni presentan una matriz de correlaciones clara, del mismo modo no rechazan la matriz de identidad conforme con la prueba de esfericidad de Bartlett, lo cual permite estimar la pertinencia para el AFE de acuerdo con Hair et al. (2014). A razón de esto, usan análisis de componentes principales (ACP) lo cual implica una combinación matemática de las variables para reducir dimensionalidad no así el estudio de constructos latentes, para lo cual se precisa un análisis factorial confirmatorio (AFC) con método de máxima verosimilitud u otro pertinente. Es decir, solo exploraron componentes lo cual no permite suponer constructos



o factores psicológicos. En este sentido, se presenta la extracción de un solo factor donde la comunalidad inicial fue fijada en 1 con comunalidades extraídas no lo suficientemente altas para sostener la existencia de un factor general fuerte. Cabe destacar que el manual no contiene ninguna justificación e interpretación del estudio. Además, en la misma lógica, mediante el método de extracción con ACP, aparece “Una estructura con cuatro factores principales que justifican el 62% de la varianza” (Vidal & González Manjón, 2019, p.27), lo cual no es coherente con las varianzas explicadas por cada factor. El cálculo correcto da como resultado 54,887%. Adicionalmente, los resultados de AFE y AFC no coinciden ni convergen matemáticamente. Las cargas factoriales se distribuyen en factores discrepantes con los índices generales los cuales, según el manual, conforman dimensiones de análisis, puntuación y otros derivados clínicos sin justificación. En detalle, las variables matemáticas cargan junto con las cognitivas y comprensión lectora, pero al mismo tiempo sin justificación forman constructos distintos. Por último, el manual solo exhibe tres índices generales: cognitivo, lectura y psicosocial, y si se recalcula en base a los resultados expuestos, se deja en manifiesto errores de cálculo que permiten inferir inflación de los puntajes o uso de distintas muestras. Para explicitar aún más esto, se revisó un grupo factorial (índice General Cognitivo), cuyos resultados están presentes en el manual 6, en la página 29. Los pesos factoriales hallados son: 0 para reflexividad, 0,84 razonamiento analógico, 0,63 para organización perceptiva y 0,29 para memoria-atención, cuya varianza explicada por variable es 0, 0,7, 0,39 y 0,08 respectivamente, cuya suma para varianza total explicada es 0,395 (39,5%), sin agregar el peso de reflexividad en la ecuación. Sin embargo, en el manual se reporta una varianza total explicada de 50,68%. Las discrepancias se observan en cada grupo factorial, denotando un error grave. También es raro un peso factorial igual a 0 en reflexividad, lo que permite sospechar sobre el uso desordenado de la información.

- D. En la página 112, el manual exhibe la escala de matemáticas y la correlación entre las variables que la componen, sean estas cálculo y numeración y resolución de problemas, cuya correlación de Pearson es 0,177, de tipo débil y cerca de ser nula en términos matemáticos. Esto es problemático ya que supone variables que deben analizarse de manera separada para la correcta interpretación, no obstante, el manual exige el uso promediado para constituir solo una



dimensión de análisis. Si se analiza la proporción de varianza del puntaje promedio atribuible a covarianza común, de acuerdo con $\frac{2r}{2+2r}$, se obtiene $\frac{0,177}{1,177} \approx 0,15$. Por lo cual, aproximadamente el 85% es ruido combinado en función de un factor unitario, más que de una dimensión latente de análisis principal. Por lo tanto, cerca del 3% de la varianza de una variable es explicada por la otra, en el marco de la constitución matemática del test.

- E. En la línea anterior, de acuerdo con lo que contiene el manual en la página 112, para el índice general matemático se usa la fórmula $IGM = CN + 3RP$, siendo CN y RP la puntuación directa del test de cálculo y numeración y resolución de problemas. Aquí se observa una fórmula sin ninguna justificación matemática ni psicométrica, lo cual es problemático puesto que multiplicar por tres el puntaje directo de resolución de problemas genera un aumento del error asociado (3 veces el error de RP), 9 veces más varianza aportada por RP y una relación donde cálculo y numeración aporta un tercio en comparación con resolución de problemas. Además, esto hace que diferencias menores en RP separe artificialmente a estudiantes, evite la homogeneidad del índice y altera la sensibilidad diagnóstica solo inflando el ruido estadístico. La fórmula de puntuación global propuesta por los autores del EVALÚA no tiene ninguna justificación psicométrica y genera alteraciones incluso diagnósticas de ser usado el instrumento para estos fines. Para hacer una simulación, a razón de la objetividad de este estudio, se usan los baremos originales que aparecen en el manual Evalúa-6: Para el test de cálculo y numeración, la media es 18,63 y la desviación estándar es 7,048; para el test de resolución de problemas la media es 22,9 y la desviación estándar es 7,834; y para la escala total, la media es 39,794 y la desviación estándar es 11,327. Entonces, si un sujeto X obtiene 3 puntos en el test de cálculo y numeración y 20 puntos en el test de resolución de problemas, estaría a más de 2 desviaciones típicas por debajo de la media respecto al primer test (cálculo y numeración) y cerca de 0 desviaciones típicas en función de la media en el segundo test (resolución de problemas), pero, aquí está el error, ya que según la fórmula que usa el test para el índice general, el resultado global del sujeto X es 2 desviaciones estándar por sobre la media, lo cual no tiene sentido y revela que el baremo presente en el manual está alterado. Si se hace el ajuste



matemático, la media debería estar cerca de 87 puntos y la desviación estándar cercana a 25 puntos, con lo cual, usando al sujeto X, da -0,96 desviaciones típicas, escondiendo entonces el evidente desarrollo bajo en cálculo y numeración. Y, al simular al sujeto B que obtuvo 18 puntos en el primer test y 3 puntos en el segundo test, se obtiene 0 desviaciones típicas con respecto a la media en la variable cálculo y -2,5 desviaciones típicas en resolución de problemas, dando como resultado (usando el baremo correcto que no coincide con lo reportado en el manual) -2,4 desviaciones típicas, lo cual revela el gran peso que tiene el segundo test, ocultando gran parte del peso de cálculo y numeración sin ninguna justificación.

- F. En la página 106 se observa que la discriminación media de los ítems del test de resolución de problemas matemáticos es de 0,209 (mínimo -0,034 y máximo 0,438), lo cual está por debajo del umbral psicométrico de ítems que diferencian entre mayor y menor dominio o desarrollo, en efecto, dichos reactivos no constituyen elementos suficientes para un test clínico y/o escolar, ya que de acuerdo con los criterios de Backhoff, Larrazolo y Rosas (2000), un reactivo con un índice de discriminación mayor de 0.39 se considera excelente; entre 0.3 y 0.39, bueno; entre 0.2 y 0.29, regular; y menor que 0.2, deficiente. Esto es grave, ya que el test con mayor peso de la dimensión matemática, es un test no apto para el estudio de habilidades y conocimientos matemáticos.

Escala Wechsler de Inteligencia para Niños-Quinta Edición

La Escala de Wechsler de Inteligencia para Niños (WISC) es una batería de pruebas formales ampliamente usada en Chile y en otros países, cuyo origen es estadounidense y ha tenido diversas adaptaciones y parametrizaciones en países de habla hispana e inglés. Desde el 2007 se usó WISC-IIIv.ch (estandarización chilena) cuya estructura factorial data del año 1991 (Rosas et al., 2022). Y, desde el año 2022, se usa la quinta versión con baremos chilenos (Rosas et al., 2022). Sin embargo, cabe destacar que WISC ha tenido presencia en Chile desde el año 1960 en su adaptación original por la Pontificia Universidad Católica de Chile (Campazzo, 1962).

WISC-V es un test de ejecución máxima de administración individual que permite la evaluación de la inteligencia en niños y adolescentes desde 6 años 0 meses hasta 16 años 11 meses. La batería presenta diversos tests que permiten medir: analogías verbales y vocabulario, constituyéndose la dimensión



Comprensión verbal; construcción de cubos y rompecabezas visuales, constituyéndose la dimensión Visoespacial; matrices de razonamiento y balanzas, constituyéndose la dimensión Razonamiento fluido; retención de dígitos y retención de imágenes, constituyéndose la dimensión Memoria de trabajo; y claves y búsqueda de símbolos, constituyéndose la dimensión Velocidad de procesamiento (Rosas et al., 2022). WISC-V es una batería de alto alcance psicométrico, lo que permite estudiar con cierto grado de profundidad el desarrollo cognitivo de niños y adolescentes. Si se analiza el trabajo de estandarización se obtiene:

- A. Se observa una metodología robusta para establecer los baremos de la batería. Los autores usaron el método de norma inferencial (Zhu y Chen, 2011), el cual asume una distribución teórica del constructo, habitualmente normal, sobre las variables latentes, y, de este modo, los parámetros no son solo descriptivos sino inferenciales, alejándose de alguna manera de la información empírica bruta de la población. Este tipo de método permite estimar puntos de corte con énfasis probabilístico mas no descriptivo absoluto, por lo cual se corrige de cierto modo las imperfecciones del proceso de obtención de información cognitiva de la población, asociado al no cumplimiento de los supuestos de medición limpia, como disposición del sujeto evaluado, errores de aplicación, fatiga u otros factores que producen error típico. Del mismo modo, usaron suavizado de curvas para garantizar continuidad con ajuste teórico y ajuste funcional de desarrollo mediante regresiones (Rosas et al., 2022).
- B. Respecto a la confiabilidad del WISC-V, los autores utilizaron el procedimiento de Split-half para la mayoría de las subpruebas y correlaciones directas entre mitades, corregido por la fórmula de Spearman-Brown. Los resultados son concluyentes respecto a la estabilidad interna de los tests, con coeficientes que van desde 0,645 a 0,941 (Rosas et al., 2022). Y, respecto al estudio mediante alfa de Cronbach y correlación de medidas repetidas, en subpruebas los resultados van desde 0,81 a 0,935.
- C. Respecto a la validez de constructo, los autores estudiaron un modelo jerárquico de cinco factores para las 10 subpruebas primarias, dando como resultado un excelente grado de ajuste (χ^2 , *RMSEA*, *CFI*, *TLI*). A nivel de subescalas y constructos, las saturaciones van desde 0,62 a



0,82, lo cual deja proporciones de varianzas no explicadas que van desde 62% a 32%; y sobre las saturaciones de las cinco dimensiones cognitivas con la escala general, se obtuvo pesos que van desde 0,65 (Velocidad de procesamiento) a 0,96 (Razonamiento fluido), lo cual deja proporciones de varianzas no explicadas que van desde 57% a 8%. Estos hallazgos permiten inferir validez de los constructos que mide la batería, sin embargo, dado el sustento teórico de las variables y la especificidad de cada test, se esperarían saturaciones sobre 0,75 para cada dimensión, lo cual no ocurre en totalidad, pero sí en la mayoría de las asociaciones.

- D. De acuerdo con el sistema de puntuación del WISC-V, el flujo consiste en calcular el puntaje bruto de cada subprueba para luego convertir a puntaje escalar (*Media* = 10; *DT* = 3), de esta manera se suman de manera simple para componer la dimensión de análisis (índice) transformándose luego en puntaje derivado tipo CI (*Media* = 100; *DT* = 15). Posteriormente, para el cálculo del cociente intelectual total (CIT), se suman de manera simple ciertos puntajes escalares para luego transformar a CI. Este modelo es simple desde el punto de vista matemático-psicométrico, y si bien es clásico y habitualmente utilizado en tests bajo la teoría clásica, genera distorsiones en la representatividad de los constructos ya que asume implícitamente una contribución equivalente de las subvariables al factor general, ignorando la diferencia en saturación de los distintos constructos y variables observables. Por ejemplo, Velocidad de procesamiento mantiene correlación de 0,65 y Razonamiento fluido de 0,96 con el factor general, llámese CIT, y, por lo tanto, razonamiento fluido aporta más del doble de varianza común que velocidad de procesamiento. Este modelo es problemático en perfiles heterogéneos, que son justamente los más clínicamente relevantes.

En este sentido, se concluye que WISC-V es una batería de amplio alcance psicométrico, fiable, válida y con consistencia teórica. Sin embargo, el sistema de puntuación $CIT \propto \text{Suma de escalares} \rightarrow \text{Conversión normativa}$, si bien es estándar y usual, no es lo psicométricamente sensible para el estudio de constructos, a diferencia de modelos más avanzados como $CIT \propto \sum \lambda_i Z_i$ (ponderación por carga factorial).



Test Básico de Aprendizaje

El Test Básico de Aprendizaje (TBA) es una batería de evaluación psicopedagógica multidimensional e informatizada, diseñada para evaluar un amplio espectro de variables cognitivas y habilidades específicas asociadas a los aprendizajes instrumentales en niños, adolescentes y adultos de 6 a 55 años (San Martín, 2025b). Esta batería fue estandarizada en Chile el año 2024 y fue masificada el año 2025 por el Centro de Estudios Psicométricos y del Aprendizaje.

El TBA es de administración individual por medio de un software y contiene diversas pruebas que permiten evaluar tasa de aprendizaje, bases del pensamiento lógico, memoria inmediata y velocidad atencional, constituyéndose la dimensión Cognitiva; bases del aprendizaje matemático, constituyéndose la dimensión Habilidades matemáticas; comprensión lectora, bases del aprendizaje lector y velocidad lectora en voz alta, constituyéndose la dimensión Lectura; y ortografía fonética y visual, constituyéndose la dimensión Ortografía. En este sentido, de acuerdo con San Martín (2025b), es un test de medición cognitiva que, al no incorporar componentes curriculares y lingüísticos en torno a lo cognitivo, presenta un sesgo socioeconómico y conceptual menor.

El TBA es una batería de alcance psicométrico que permite estudiar el desarrollo cognitivo y las habilidades académicas de las personas. Si se analizan sus propiedades psicométricas se obtiene:

- A. En cuanto a su fiabilidad, el autor utilizó alfa de Cronbach para saber sobre la consistencia interna de cada test integrado en la batería, cuyo promedio, considerando todas las variables es de 0,77, lo que sugiere una consistencia interna de moderada a alta de acuerdo con Celina Oviedo y Campo Arias (2005) (San Martín, 2025b). Para la fiabilidad temporal, se usó el método test-retest a través de la correlación de Pearson y un modelo de regresión lineal en un intervalo de dos semanas, lo cual arrojó una asociación fuerte superior a 0,9.
- B. Respecto a la validez de constructo, el trabajo presenta diversos análisis de rigor, estos son: adecuación de los datos mediante el índice de Kaiser-Meyer-Olkin (KMO), prueba de esfericidad de Bartlett, siguiendo los criterios establecidos por Garrido Bermúdez et al. (2023), índices de ajuste como CFI, TLI, RMSEA y SRMR, método de máxima verosimilitud y rotaciones como Varimax (ortogonal), de modo tal que se justifique el análisis factorial



exploratorio y confirmatorio (San Martín, 2025b). En general, los resultados son favorables dada la naturaleza multi-habilidad de cada test, con saturaciones que van desde 0,55 (Velocidad atencional) a 0,67 (Bases del pensamiento lógico) en la dimensión cognitiva, dejando varianzas no explicadas que van desde 69,7% a 55%; y en habilidades académicas, las saturaciones van desde 0,53 (Bases del aprendizaje lector) a 0,83 (Comprensión lectora), con varianzas no explicadas de 72% a 31%.

- C. De acuerdo con el sistema de puntuación, el flujo consiste en calcular el puntaje directo (PD) mediante diversas fórmulas según cada test, así como la media de la población correspondiente al grupo etario y a la experiencia escolar, junto con la desviación estándar de cada variable. Posteriormente, se calcula la puntuación tipificada para determinar la posición en la distribución normal. Para la estimación del desarrollo cognitivo bajo un único factor, se consideran BPL, MI y A, calculando el Cociente Intelectual (CI) con una media de 100 y una desviación estándar de 15, así como el puntaje T con una media de 50 y una desviación estándar de 10; a este proceso se le denomina síntesis psicométrica. Finalmente, para la estimación del cambio en las variables, se utiliza el cálculo basado en T, aplicando la ecuación $\frac{T_2 - T_1}{T_1}$ en relación con los indicadores globales de cognición, lectura, matemáticas y ortografía (San Martín, 2025b).

En este sentido, se concluye que TBA es una batería de alcance psicométrico, fiable, válida y con consistencia teórica. Sin embargo, el sistema de puntuación con aritmética simple, si bien es estándar y usual, no es lo psicométricamente sensible para el estudio de constructos, a diferencia de modelos más avanzados como $X \propto \sum \lambda_i Z_i$ (ponderación por carga factorial).

Descripción del Test Básico de Aprendizaje Profundo

El Test Básico de Aprendizaje Profundo (TBA-P) es una batería de evaluación psicopedagógica multidimensional de medición profunda, estandarizada e informatizada, diseñada para evaluar el aprendizaje humano considerando su complejidad y, por lo tanto, su aproximación a las aptitudes académicas y al desarrollo cognitivo. El instrumento se estructura sobre 11 pruebas unidimensionales, donde cada prueba posee una lógica específica de medición y condiciones estandarizadas de aplicación.

El TBA-P se implementa a través del software TBA, el cual incorpora algoritmos psicométricos



avanzados, corrección automática e informes interpretativos globales y específicos por grupo de variables. Un elemento central del sistema es la duración dinámica en función del control techo, la cual se operacionaliza mediante un algoritmo de control basado en el cálculo de error estable, permitiendo interrumpir la administración de cada prueba cuando la probabilidad de información adicional significativa se vuelve irrelevante. Este método genera mayor eficiencia, menor efecto azar, control de fatiga y estabilidad en la estimación del rendimiento, manteniendo la validez de la medición cognitiva (International Test Commission & Association of Test Publishers, 2025).

Desde el punto de vista conceptual, el TBA-P distingue explícitamente entre aprendizajes primitivos, aprendizajes instrumentales y procesos cognitivos, evitando la confusión entre carga curricular y habilidades intelectuales subyacentes. En este marco, las variables cálculo mental y razonamiento matemático son tratadas como universales, en tanto constituyen aprendizajes primitivos no dependientes del currículo formal, sino de la capacidad básica para operar mentalmente con cantidades y resolver problemas matemáticos cotidianos sin apoyo de instrumentos externos; en cuanto a habilidades numéricas, comprensión lectora, ortografía y velocidad lectora, constituyen aprendizajes instrumentales más susceptibles a la experiencia escolar; y atención, memoria e inteligencia, constituyen procesos cognitivos de básicos-ejecutivos a superiores.

Tabla 1. Objetivos específicos del TBA-P.

Test	Objetivo
Cálculo mental	Medir el procesamiento cognitivo para sumar y restar mentalmente sin apoyo de instrumentos externos.
Razonamiento matemático	Medir el razonamiento matemático a través de problemas prácticos sin apoyo de instrumentos externos.
Habilidades numéricas	Medir el procesamiento cognitivo para manipular números y representaciones numéricas sin apoyo de instrumentos externos.
Comprensión lectora	Medir la comprensión lectora a través de textos breves.
Velocidad lectora de palabras habituales	Medir la velocidad lectora de palabras habituales a través de un texto breve.
Velocidad lectora de pseudopalabras	Medir la velocidad lectora de pseudopalabras a través de un texto breve.

pseudopalabras	
Ortografía visual	Medir el desarrollo de esquemas ortográficos para el reconocimiento de la escritura reglada.
Ortografía fonética	Medir el desarrollo de esquemas fonológicos-grafémicos para la conversión de estructuras fonéticas en palabras escritas.
Inteligencia general	Medir la inteligencia general a través de la percepción de estímulos geométricos visuales que exigen educación de correlatos y relaciones.
Atención- Concentración	Medir la capacidad de concentración a través de la asociación de figuras en un plano gráfico.
Memoria de corto plazo	Medir la memorización de figuras en un plano gráfico.

Elaboración propia.

Método para el estudio de un test psicométrico-psicopedagógico de alta precisión denominado Test Básico de Aprendizaje Profundo

El estudio se desarrolló bajo un enfoque psicométrico cuantitativo, con un diseño no experimental, transversal e instrumental, orientado al análisis de la estructura interna y la consistencia de medición del TBA-P, trabajando a nivel de puntajes agregados por dominio para el estudio de modelos cognitivos y estructurales. La muestra estuvo compuesta por 178 participantes (6 a 55 años), con distribución equilibrada por sexo y procedentes de diversas regiones de Chile; para los análisis confirmatorios se utilizaron tamaños muestrales efectivos de $N = 160$ y $N = 145$. Los datos provinieron del software TBA-P, que administra los tests de forma informatizada con control algorítmico del proceso de aplicación; cada variable correspondió a un puntaje derivado de ítems dicotómicos pertenecientes a tests unidimensionales, organizados en aprendizajes primitivos, aprendizajes instrumentales y procesos cognitivos generales. El procesamiento inicial se realizó en Microsoft Excel, donde se estimó la consistencia interna mediante KR-20, y los análisis factoriales se llevaron a cabo en RStudio utilizando los paquetes psych y lavaan; la adecuación factorial se evaluó mediante KMO y Bartlett, el AFE se estimó mediante ejes principales (PA) con rotación oblimin, y el AFC mediante máxima verosimilitud

robusta (MLR), evaluando el ajuste a través de CFI, TLI, SRMR y RMSEA, interpretando este último con cuidado dadas las condiciones de tamaño muestral y grados de libertad.

Evidencia de fiabilidad a través de la consistencia interna

La consistencia interna de cada uno de los tests fue estimada mediante el coeficiente Kuder–Richardson 20 (KR-20), dado que los ítems presentan formato de respuesta dicotómica. Este coeficiente es matemáticamente equivalente al alfa de Cronbach cuando las respuestas son binarias, por lo que constituye un estimador apropiado de la fiabilidad en este contexto. Los coeficientes de consistencia interna obtenidos fueron los siguientes:

Tabla 2. Coeficientes de consistencia interna de los tests integrados en la batería TBA-P.

Test	Resultado KR-20
Inteligencia general	0,93
Atención–Concentración	0,86
Memoria de corto plazo	0,85
Habilidades numéricas	0,92
Cálculo mental	0,95
Razonamiento matemático	0,96
Comprensión lectora	0,96
Ortografía visual	0,97
Ortografía fonética	0,96

Elaboración propia.

En conjunto, los resultados evidencian niveles elevados de consistencia interna en todos los tests, con valores que superan ampliamente el umbral mínimo aceptable de 0,7 y, en todos los casos, se sitúan en rangos considerados óptimos para instrumentos de evaluación psicológica y psicopedagógica (Celina Oviedo y Campo Arias, 2005).

Si bien varios coeficientes superan el valor de 0,9, este comportamiento es coherente con la estructura unidimensional de cada prueba, diseñada para evaluar un único proceso o dominio cognitivo específico.

En este contexto, los altos niveles de consistencia interna no se interpretan como redundancia espuria,



sino como consecuencia de la homogeneidad conceptual y funcional de los ítems que componen cada instrumento.

Evidencia de validez del entorno cognitivo del TBA-P

Previo a la extracción factorial, se evaluó la adecuación de la matriz de correlaciones entre las variables Inteligencia general (IG), Atención-Concentración (A), Memoria de corto plazo (M), Cálculo mental (CM) y Razonamiento matemático (RM). El índice Kaiser–Meyer–Olkin (KMO) obtuvo un valor global de 0,78, lo que indica una adecuación muestral buena para la aplicación de análisis factorial. A nivel individual, los valores MSA fluctuaron entre 0,72 y 0,91, todos por sobre el umbral mínimo recomendado, lo que sugiere que las variables comparten varianza suficiente para ser factorizadas. De acuerdo con Garrido Bermúdez et al. (2023), valores de KMO superiores a 0,70 son indicativos de matrices factorialmente estables.

Adicionalmente, la prueba de esfericidad de Bartlett resultó estadísticamente significativa ($\chi^2 = 592,88$; $gl = 10$; $p < 0,001$), rechazándose la hipótesis nula de matriz identidad y confirmando la presencia de correlaciones suficientes entre las variables para continuar con el análisis factorial (Garrido Bermúdez et al., 2023).

Tabla 3. Matriz de correlaciones entre variables cognitivas.

Variable	IG	A	M	CM	RM
IG	1.000	0.534	0.578	0.686	0.710
A	0.534	1.000	0.848	0.543	0.469
M	0.578	0.848	1.000	0.597	0.541
CM	0.686	0.543	0.597	1.000	0.838
RM	0.710	0.469	0.541	0.838	1.000

Elaboración propia mediante código en RStudio.

En este sentido, se realizó un análisis factorial exploratorio mediante ejes principales (PA) con rotación oblimin, considerando la posible correlación entre factores. La solución de dos factores mostró una estructura conceptualmente interpretable. El Factor 1 agrupó a IG, CM y RM, con cargas factoriales

altas ($\geq 0,63$), mientras que el Factor 2 agrupó a A y M, con cargas muy altas ($\geq 0,88$). Esta solución explicó conjuntamente el 76,4 % de la varianza total, lo que evidencia una estructura latente consistente, compuesta por dominios cognitivos diferenciables pero relacionados.

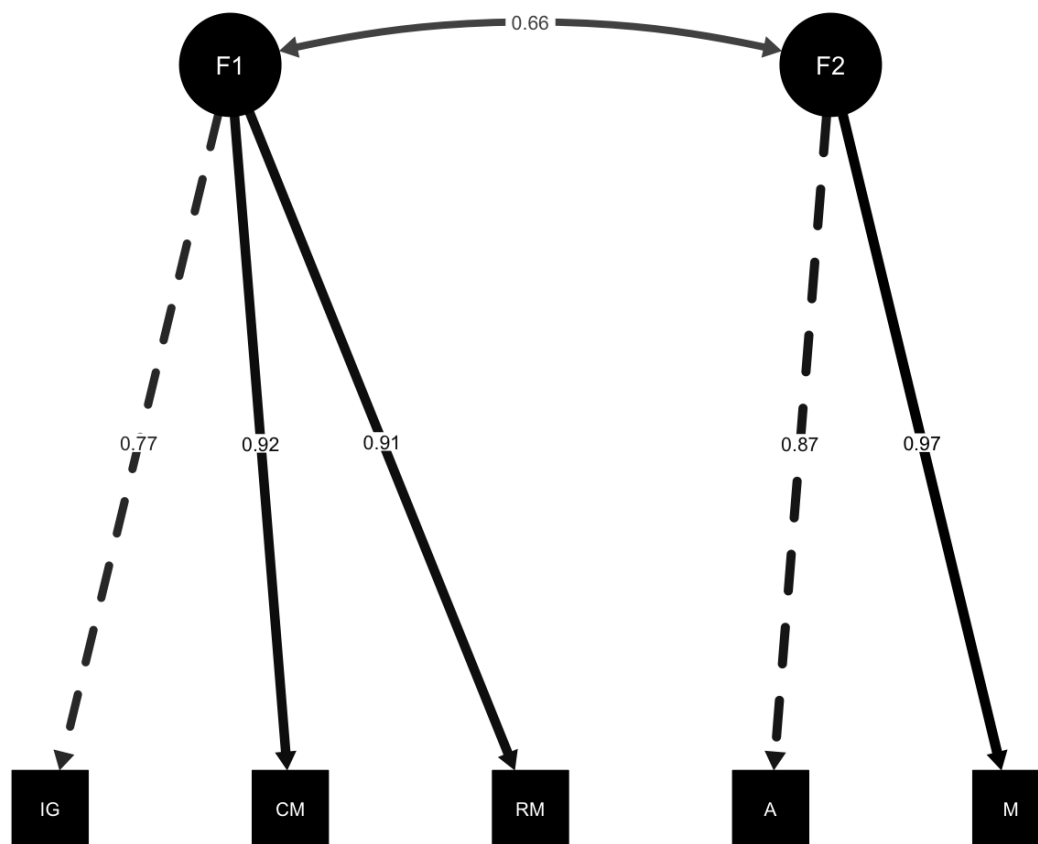
Por último, se especificó un modelo factorial confirmatorio de dos factores correlacionados, donde el primer factor se definió por IG, CM y RM, y el segundo por A y M.

Los resultados mostraron índices de ajuste global adecuados: CFI = 0,986, TLI = 0,965 y SRMR = 0,037, todos dentro de los rangos considerados aceptables en la literatura especializada. Según la revisión de Jordan Muiños (2021), valores de CFI y TLI iguales o superiores a 0,90 indican un ajuste adecuado y valores superiores a 0,95 revelan un ajuste óptimo. Del mismo modo, en muestras superiores a 100 casos, el SRMR inferior a 0,08 se interpreta como evidencia de una adecuada reproducción de la matriz de covarianzas observada (Jordan Muiños, 2021).

El estadístico χ^2 resultó estadísticamente significativo ($\chi^2 = 11,167$; gl = 4; p = 0,025), lo cual es esperable considerando la sensibilidad de este índice al tamaño muestral y a modelos con bajo número de grados de libertad (Jordan Muiños, 2021).

En cuanto a los parámetros estimados, todas las cargas factoriales estandarizadas fueron altas y estadísticamente significativas, con valores que fluctuaron entre 0,77 y 0,97, lo que evidencia una fuerte relación entre las variables observadas y sus respectivos factores latentes. La correlación entre los dos factores fue moderada–alta ($r = 0,66$), lo que supone la presencia de una base cognitiva común subyacente, sin perder la diferenciación funcional entre los dominios cognitivos representados.

Figura 1. Modelo cognitivo bifactorial.



Elaboración propia mediante código en RStudio.

Evidencia de validez del entorno lecto-ortográfico del TBA-P

En primer lugar, la adecuación de los datos para análisis factorial fue favorable. El índice KMO arrojó un valor global $MSA = 0,81$, con valores por variable entre 0,75 y 0,85, lo que indica correlaciones suficientemente altas y, por tanto, una matriz apta para ser factorizada. En complemento, la prueba de esfericidad de Bartlett fue significativa ($\chi^2 = 480,79$; $gl = 10$; $p < 0,001$), lo que rechaza la hipótesis de matriz identidad y confirma la presencia de covarianzas suficientes para justificar la extracción factorial (Garrido Bermúdez et al., 2023).

Tabla 4. Matriz de correlaciones entre variables lecto-ortográficas.

Variable	CL	OV	OF	VLPH	VLPS
CL	1.000	0.723	0.749	0.602	0.614
OV	0.723	1.000	0.725	0.494	0.461
OF	0.749	0.725	1.000	0.588	0.566
VLPH	0.602	0.494	0.588	1.000	0.807
VLPS	0.614	0.461	0.566	0.807	1.000

Elaboración propia mediante código en RStudio.

Con estos supuestos verificados, el análisis factorial exploratorio por ejes principales (PA) con rotación oblimin mostró una solución de dos factores. El Factor 1 concentró a Comprensión lectora (CL), Ortografía visual (OV) y Ortografía fonética (OF) con cargas altas (0,760; 0,925; 0,804), mientras que el Factor 2 concentró a Velocidad lectora de palabras habituales (VLPH) y Velocidad lectora de pseudopalabras (VLPS) con cargas altas (0,851; 0,922). La varianza explicada acumulada fue 74%, con una contribución de 41,6% para el primer factor y 32,3% para el segundo, lo que sugiere una estructura latente robusta y parsimoniosa en términos de dominios diferenciables.

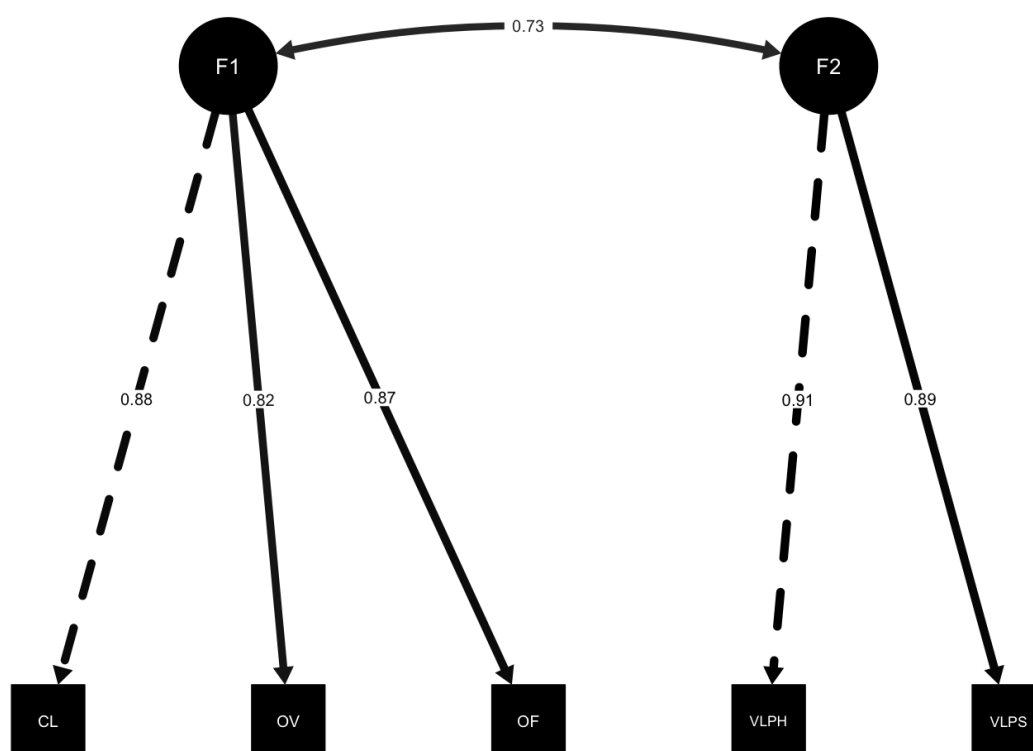
A razón de esta estructura, el análisis factorial confirmatorio produjo un modelo de dos factores correlacionados (F1: CL, OV, OF; F2: VLPH, VLPS) ($r=0,73$), con saturaciones que van desde 0,82 a 0,91. Los índices incrementales evidenciaron ajuste óptimo: CFI = 0,993 (robusto = 1,000) y TLI = 0,982 (robusto = 1,007). Conforme con Jordan Muiños (2021), valores de CFI $\geq 0,95$ se interpretan como ajuste del modelo a la muestra y TLI $> 0,90$ como criterio de ajuste adecuado.

En la misma línea, el SRMR = 0,026 indicó un ajuste estrecho entre la matriz de covarianzas observada y la reproducida por el modelo. Este índice ha sido destacado como particularmente eficaz para discriminar modelos con mal ajuste en muestras pequeñas, especialmente en rangos cercanos a los 200 participantes, debido a su menor sensibilidad al tamaño muestral y a los grados de libertad del modelo (Jordan Muiños, 2021). Además, el RMSEA = 0,077 (estimación no robusta) debe interpretarse con cuidado, debido a que este índice presenta limitaciones importantes en muestras inferiores a 200 casos,

recomendándose explícitamente su uso con precaución en contextos muestrales menores, ya que tiende a sobreestimar el error de aproximación en modelos con bajo número de grados de libertad. Esta condición se cumple en el presente estudio ($N = 145$; $gl = 4$), lo que justifica una interpretación crítica del RMSEA y su consideración integrada con otros índices de ajuste (Jordan Muiños, 2021).

Respecto del estadístico χ^2 , el resultado no fue estadísticamente significativo en su versión escalada ($p = 0,496$), lo que es consistente con un ajuste adecuado del modelo. No obstante, Jordan Muiños (2021) advierte que este estadístico es altamente sensible al tamaño muestral, pudiendo conducir tanto al rechazo como a la aceptación del modelo por razones ajenas a su adecuación sustantiva. En efecto, su interpretación debe realizarse siempre en conjunto con índices incrementales y absolutos complementarios.

Figura 2. Modelo lecto-ortográfico bifactorial.



Elaboración propia mediante código en RStudio.

Modelo de puntuación TBA-P

El sistema de puntuación del TBA-P se fundamenta en un enfoque factorial-estructural, cuyo objetivo es estimar una variable latente de inteligencia general a partir de procesos cognitivos-ejecutivos y de aprendizajes primitivos, manteniendo simultáneamente la posibilidad de analizar cada dominio de manera individual y no latente.

La Inteligencia factorial (IF) se concibe como un constructo latente que resume la covariación entre cinco variables observadas: IG, CM, RM, A y M. Previo al cálculo factorial, todas las variables son transformadas a puntajes tipificados (Z), de acuerdo con las normas correspondientes por edad, con el fin de operar sobre una métrica común.

La estructura factorial de referencia identifica dos dimensiones de primer orden: una dimensión cognitiva-matemática, integrada por IG, CM y RM, y una dimensión ejecutiva, compuesta por A y M. Las cargas factoriales estandarizadas obtenidas del modelo estructural se utilizan como pesos para la estimación del constructo. En efecto, la IF se calcula como un promedio ponderado de los puntajes Z de las variables observadas, según la expresión general:

$$Z_{IF,i} = \frac{\sum_{j=1}^k \lambda_j Z_{ij}}{\sum_{j=1}^k \lambda_j}$$

En el caso de la evidencia del TBA-P, la estimación específica es:

$$Z_{IF,i} = \frac{0,77Z_{IG} + 0,92Z_{CM} + 0,91Z_{RM} + 0,87Z_A + 0,97Z_M}{4,44}$$

El valor resultante corresponde a una estimación operativa del factor general de inteligencia, expresada en una escala tipificada con media cero y desviación estándar uno.

En conformidad, cada una de las variables que conforman la IF es analizada de manera individual en su forma observada, mediante la interpretación de sus puntajes brutos y tipificados. Este procedimiento responde al principio de que el análisis factorial explica la varianza compartida entre procesos, pero no sustituye la evaluación específica de los procesos cognitivos subyacentes, los cuales conservan relevancia diagnóstica y descriptiva propia.

La misma lógica se usa para las variables CL, OV, OF, VLPH y VLPS, las cuales conforman dos factores lectores correlacionados que constituyen el Desarrollo lector (DL): semántico-ortográfico-fonológico y



velocidad de procesamiento lector por vías visuales y fonológicas. El modelo es:

$$Z_{DL,i} = \frac{\sum_{j=1}^k \lambda_j Z_{ij}}{\sum_{j=1}^k \lambda_j}$$

En el caso de la evidencia del TBA-P, la estimación específica es:

$$Z_{DL,i} = \frac{0,88Z_{CL} + 0,82Z_{OV} + 0,87Z_{OF} + 0,91Z_{VLPH} + 0,89Z_{VLPS}}{4,37}$$

El valor resultante corresponde a una estimación operativa del factor de lectura, expresada en una escala tipificada con media cero y desviación estándar uno.

CONCLUSIÓN

Los resultados del presente trabajo permiten concluir que la evaluación psicopedagógica de alta precisión exige, de manera ineludible, coherencia epistemológica, rigor matemático y consistencia psicométrica entre el modelo teórico, el diseño del instrumento y el sistema de puntuación utilizado. El análisis crítico de baterías ampliamente difundidas muestra que la estandarización formal y la masividad de uso no garantizan, por sí mismas, validez interpretativa ni sensibilidad diagnóstica, especialmente cuando los modelos de síntesis ignoran la estructura latente real de los constructos evaluados o incorporan procedimientos aritméticos carentes de justificación psicométrica.

En este contexto, el TBA-P se posiciona como una alternativa metodológicamente consistente, al articular una distinción explícita entre aprendizajes primitivos, aprendizajes instrumentales y procesos cognitivos, evitando la confusión entre rendimiento curricular y habilidades subyacentes. La evidencia presentada respalda la fiabilidad de sus tests unidimensionales, la validez factorial de sus entornos cognitivo-ejecutivo y lecto-ortográfico, y la pertinencia de un modelo de puntuación basado en ponderación factorial, el cual permite estimar variables latentes sin perder la especificidad de los procesos observados.

Desde una óptica psicométrica y clínica, el diseño del TBA-P aporta una solución concreta a problemas históricos de la evaluación del aprendizaje, particularmente en el estudio de trastornos específicos del aprendizaje, donde la precisión diagnóstica depende de la correcta separación entre déficit estructural, variabilidad cognitiva y efectos del contexto educativo (American Psychiatric Association, 2014; Escobar & Tenorio, 2022). En consecuencia, el modelo propuesto no solo



representa un avance técnico en términos de medición, sino que establece un marco conceptual para el desarrollo de instrumentos psicopedagógicos orientados a una evaluación válida, fiable y clínicamente significativa.

Financiamiento

El presente trabajo fue financiado con fondos privados a través de la empresa Centro de Estudios Psicométricos y del Aprendizaje SpA.

Agradecimientos

El autor de este trabajo agradece especialmente a las y los profesionales que participaron como examinadores en el trabajo de campo, cuya labor fue fundamental para la correcta aplicación de los instrumentos y la recolección rigurosa de la información empírica proveniente de la población evaluada.

En este proceso participaron Valeria Amor de los Ángeles Escobar Ibacache, Marcela Andrea López Osorio, Paulina Alejandra Vargas Araya, Belén Josefa Alejandra Caamaño Salort, Valentina Esperanza González Catalán, Karla Isabel Gaete Pavez, Nicol Elizabeth Neira Torres, Sandra del Carmen Franco Cornejo, Alejandra del Carmen Olea Ossa, Matías Alejandro Ortega Santibáñez, Lizardo Andrés Castro Gamonal, Daniela Roxana Marín Huerta, Alicia Soledad Cáceres Echeverría, Claudia Patricia Gómez Guerra, María José Munizaga Pizarro, Niel Andrés Aravena Vera, María Cecilia Campos Castro, Montserrat Mariapia Leal Lisboa, Carolina Pamela Gahn Basaure, Verónica Elizabeth Martínez San Martín, Nicole Andrea Muñoz Conejeros, Verónica Lynn Zanetti Barra y Evelyn Martínez San Martín.

REFERENCIAS BIBLIOGRÁFICAS

- American Psychiatric Association. (2014). *DSM-5: Manual diagnóstico y estadístico de los trastornos mentales*.
- Backhoff, E., Larrazolo, N., & Rosas, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de Investigación Educativa*, 2(1). <http://redie.uabc.mx/vol2no1/contenido-backhoff.html>
- Campazzo, E. (1962). *Estandarización de la escala de Wechsler para la medición de la inteligencia en los niños (WISC) [Memoria de grado no publicada]*. Pontificia Universidad Católica de Chile.



- Celina Oviedo, H., & Campo Arias, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34(4), 572–580. <https://www.redalyc.org/pdf/806/80634409.pdf>
- Escobar, J. P., & Tenorio, M. (2022). Trastornos específicos del aprendizaje: Origen, identificación y acompañamiento. *Revista Médica Clínica Las Condes*, 33(5), 473–479.
- Garrido Bermúdez, E., Mena Rodríguez, H. Y., Zuluaga Arango, J. M., & Pérez Quintero, F. E. (2023). Proceso para validar un instrumento de investigación por medio de un análisis factorial. *UNACIENCIA*, 16(30), 61–73. <https://doi.org/10.35997/unaciencia.v16i30.724>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson Education Limited.
- International Test Commission & Association of Test Publishers. (2025). *Guidelines for technology-based assessment*.
- Jordan Muiños, F. M. (2021). Valor de corte de los índices de ajuste en el análisis factorial confirmatorio. *PSOCIAL*, 7(1), 66–78. <https://portal.amelica.org/ameli/journal/123/1232225009/>
- Muñiz, J. (2018). *Introducción a la psicometría: Teoría clásica y TRI*. Ediciones Pirámide.
- Rosas, R., Pizarro, M., Grez, O., Navarro, V., Tapia, D., Arancibia, S., Muñoz-Quezada, M. T., Lucero, B., Pérez-Salas, C. P., Oliva, K., Vizcarra, B., Rodríguez-Cancino, M., & von Fredeen, P. (2022). Estandarización chilena de la Escala Wechsler de Inteligencia para Niños–Quinta Edición. *Psykhé*, 31(1). <https://doi.org/10.7764/psykhe.2020.21793>
- San Martín Catalán, P. J. (2025a). *La lógica detrás del Test Básico de Aprendizaje: Problemas y soluciones en la evaluación psicopedagógica*. Ediciones CEPALONLINE.
- San Martín Catalán, P. (2025b). Diseño, confiabilidad y validez del Test Básico de Aprendizaje. *Revista Neuronum*, 11(3), 1-20. <https://dialnet.unirioja.es/servlet/articulo?codigo=10034258>
- Vidal, J. G., & González Manjón, D. (2019). *Batería psicopedagógica Evalúa 6: Manual* (G. López Devillaine & P. Siña Fernández, Adapt.; J. G. Vidal, Coord.) (Versión chilena 4.0). Instituto de Evaluación Psicopedagógica EOS, S.A.
- Zhu, J., & Chen, H.-Y. (2011). Utility of inferential norming with smaller sample sizes. *Journal of Psychoeducational Assessment*, 29(6), 570–580. <https://doi.org/10.1177/0734282910396323>

