

Propiedades psicométricas y estandarización de la subprueba de cálculo matemático del WRAT-4 en adolescentes colombianos

Psychometric Properties and Standardization of the WRAT-4 Mathematical Subtest in Colombian Adolescents

Recibido: febrero 14/2023; Concepto de evaluación: julio 22/2024; Aceptado: julio 23/2024

Javier Humberto Parra Pulido¹

Universidad Autónoma de Bucaramanga, Yopal, Colombia
ORCID: <https://orcid.org/0000-0003-1632-2824>

Resumen

El objetivo de este estudio fue obtener las propiedades psicométricas de la forma B de la subprueba de cálculo matemático del WRAT-4 (*Wide Range Achievement Test*) en adolescentes colombianos. Para ello, se contó con la participación de 722 estudiantes de educación básica de los grados noveno, décimo y once de instituciones públicas y privadas en Colombia. Las propiedades psicométricas se evaluaron a partir del análisis de ítems (índices de discriminación y dificultad), su validez de constructo (análisis factorial exploratorio y confirmatorio), confiabilidad (KR-20 y omega de McDonald's), y la obtención de puntuaciones normativas, con el análisis de efectos, y percentiles. Los resultados mostraron que el subtest presenta índices adecuados de discriminación ($> .30$) y distribución de los ítems en dificultad $\in (.55, .74)$, así como en su validez de contenido (CMIN/DF = 1.83, GFI = .93, CFI = .92, TLI = .91, RMSEA = .03, AIC = 774.38) y confiabilidad (KR-20 = .81, $\omega = .71$). En conclusión, la subprueba matemática del WRAT-4 puede ser utilizada para la evaluación de las habilidades matemáticas de aritmética, fracciones y álgebra en adolescentes colombianos.

Palabras clave

matemáticas, validación, análisis factorial, adolescencia.

Abstract

The aim of this study was to obtain the psychometric properties of Form B of the WRAT-4 (Wide Range Achievement Test) mathematical calculation subtest in Colombian adolescents. The study included the participation of 722 students from ninth, tenth, and eleventh grades in public and private schools in Colombia. The psychometric properties were evaluated through item analysis (discrimination and difficulty indices), construct validity (exploratory and confirmatory factor analysis), reliability (KR-20 and McDonald's omega), and the generation of normative scores, with effect size analysis and percentiles. The results showed that the subtest presented adequate discrimination indices ($> .30$) and item difficulty distribution $\in (.55, .74)$, as well as content validity (CMIN/DF=1.83, GFI = .93, CFI = .92, TLI = .91, RMSEA = .03, AIC = 774.38) and reliability (KR-20 = .81, $\omega = .71$). In conclusion, the WRAT-4 mathematical subtest can be used to assess arithmetic, fractions, and algebra skills in Colombian adolescents.

Keywords

mathematics, validity, factor analysis, high school.

Cómo citar [APA]:

Parra, J. H. (2024). Propiedades psicométricas y estandarización de la subprueba de cálculo matemático del WRAT-4 en adolescentes colombianos. *Acta Colombiana de Psicología*, 27(2), 111-129. <https://doi.org/10.14718/ACP.2024.27.2.7>

.....

1 **Datos de contacto:** Kilómetro 2 vía Matepantano, Campus Unisangil, Yopal. Correo electrónico: jparra@unisangil.edu.co

Nota del autor: Artículo resultado del proyecto de investigación producto de la convocatoria "Gestión de la producción intelectual, traspasando fronteras del conocimiento" (F-GDO-015-007), Unisangil. Agradecimientos: a los colegios y participantes de la presente investigación, así como a los estudiantes del semillero de investigación de Neuropsicología Básica y Aplicada (Neurobab) del grupo de investigación en Ciencias de la Educación y la Salud (ICES, Unisangil) por la asistencia técnica en la aplicación de las pruebas. Para acceder al instrumento validado, comunicarse al correo del autor.

Conflicto de intereses: El autor no refiere conflictos de intereses.

Introducción

Las *habilidades matemáticas* son un conjunto de capacidades que permiten identificar, entender, planear, desarrollar y resolver ideas matemáticas de forma adecuada (Gilmore et al., 2018), que implican la interpretación de los datos numéricos y la manipulación de información previa con el fin de establecer relaciones, analizar regularidades, establecer patrones de cambio, y realizar procedimientos y juicios lógicos para resolver un problema (Butterworth, 2010).

La adquisición de estas habilidades comienza justo antes de la escolarización, etapa en la cual el niño posee una representación matemática no simbólica donde procesa la información numérica, pero sin que tenga necesariamente un conocimiento previo del símbolo numérico (Kroesbergen et al., 2009; Le Corre et al., 2006). Al iniciar la escolarización, los niños adquieren una habilidad numérica simbólica que incluye la representación mental abstracta del número, incluida la escritura de las cantidades numéricas y símbolos arábigos que representan numerosidad (Butterworth, 2010).

La habilidad matemática se desarrolla a partir de la escolarización debido a que es allí donde se aprende a realizar operaciones aritméticas básicas de sumas y restas de cantidad creciente, para posteriormente llevar a cabo operaciones de mayor demanda cognitiva, como las fracciones o el álgebra, hasta llegar a realizar operaciones de mayor complejidad, como el cálculo o la trigonometría, en las etapas más avanzadas del desarrollo en edad adolescente y en la adultez (Artemenko et al., 2018; Looi et al., 2016). Teniendo esto en cuenta, se puede decir que el proceso de adquisición de habilidades matemáticas tiene una alta demanda de recursos cognitivos, ya que implica realizar tareas complejas que requieren habilidades de deducción y comprensión, debido a la representación abstracta que supone la solución de un problema matemático.

A partir de esto, la *neurociencia* ha estudiado la actividad cerebral implicada en la realización de problemas aritméticos, y ha identificado, por dar un ejemplo, el *modelo de triple código* (Arsalidou & Taylor, 2011), según el cual para la resolución de problemas matemáticos no solo se requiere de la actividad de los niveles de procesamiento visual normal, sino también de un análisis semántico-verbal que se da con la activación de distintas regiones cerebrales.

Específicamente, en estudios de *neuroimagen funcional (RMf)* se ha identificado la activación cerebral por medio de señales *Blood Oxygen Level Dependent (BOLD)* al momento de realizar operaciones matemáticas en áreas de asociación occipito-temporal ventral inferior, regiones de la corteza parietal inferior, y áreas perisilvanas del hemisferio cerebral izquierdo; todas encargadas de diversos niveles de procesamiento (Artemenko et al., 2018; Klein et al., 2013; Looi et al., 2016).

Incluso, se ha encontrado un bajo nivel de concentración del neurotransmisor GABA en el giro frontal medial del encéfalo y en la conectividad frontoparietal en estudiantes adolescentes que decidieron no cursar matemáticas en su educación secundaria —debido a la opción de elección curricular que tienen los estudiantes en países como Reino Unido— (Zacharopoulos et al., 2021); neurotransmisor que se ha asociado a la formación de nuevas redes neuronales y a la plasticidad cerebral (Frangou et al., 2018). Por esta razón, se podría identificar a la habilidad matemática como con un papel importante en el desarrollo de los adolescentes a nivel cognitivo.

Asimismo, el estudio de la *cognición matemática* ha mostrado la importancia de los procesos matemáticos en todas las edades del desarrollo, sobre todo por su relación con el éxito académico y laboral a largo plazo (para una revisión de la importancia de las matemáticas durante el ciclo vital, véase Silver & Libertus, 2022).

No obstante, a pesar de la importancia de estos hallazgos, lo cierto es que existen diversas dificultades a nivel matemático a nivel mundial tanto en los años de escolarización como después de este periodo, pues, como se encuentra en un estudio llevado a cabo en 24 países por la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2018), más del 20 % de los adultos presentaron niveles bajos en matemáticas, con dificultades importantes a la hora de realizar operaciones de dos pasos y entender el uso de decimales, porcentajes y fracciones.

Estas dificultades también se presentan en los años de escolarización, y, en particular, en Colombia se ha encontrado que el promedio de competencia matemática de los jóvenes adolescentes es *bajo* en comparación con varios países de la región y con el promedio mundial. Según los resultados de las pruebas PISA del

2018, el promedio de competencia matemática de los adolescentes colombianos es de 391 puntos, lo cual los ubica en el *nivel 1*, correspondiente a un desempeño *inferior* (OCDE, 2018). De igual forma, la prevalencia de problemas del aprendizaje matemático se ubica entre el 3.4 % y el 7.8 % tanto en niños como en adolescentes (Dorneles, 2019), aunque la variabilidad en estos resultados puede deberse al género, las características específicas de la población, e incluso a los puntos de corte utilizados en los distintos estudios (Devine et al., 2013; Reigosa-Crespo et al., 2012). No obstante, a pesar de la importancia de las matemáticas a lo largo del ciclo vital y los índices de problemas asociados a este constructo, son pocos los test que evalúan estas habilidades en adolescentes en Colombia.

El uso de *instrumentos psicométricos* es fundamental para los procesos de evaluación psicológica, ya que aportan valores estandarizados del desarrollo, desempeño, diagnóstico y rehabilitación de los distintos fenómenos psicológicos (Hebben & Milberg, 2011), pero, aunque el uso de material psicotécnico ha venido en ascenso, sus aplicaciones no cumplen con los criterios de validez esperados, ya que en su mayoría fueron desarrollados en un contexto diferente al país donde se evalúan. Al respecto, es importante tener en cuenta que el contexto sociocultural y el lenguaje tienen un efecto en la forma en la que se evalúan, estandarizan e interpretan los test (Cohen & Swerdlik, 2001; Herrera & León, 2022), y que, si un test no va dirigido ni se ha adecuado a una población particular con características culturales propias, se podría decir que el constructo no se está representando de la mejor manera (Argibay, 2006).

Adicional a esto, los profesionales suelen emplear datos normativos de otras poblaciones, hacer procedimientos personalizados, realizar cambios en las pruebas, o trabajar con puntajes brutos sin usar datos normativos que permitan obtener una valoración confiable (Olabarrieta-Landa et al., 2016), lo cual puede llevar a que se generen diagnósticos sesgados si no se evalúan todas las habilidades —matemáticas, para este caso— en su totalidad.

De hecho, se ha identificado que una de las mayores necesidades del ejercicio de la neuropsicología en el contexto latinoamericano se relaciona con la falta de acceso a pruebas y recursos profesionales para la correcta evaluación y diagnóstico (Arango-Lasprilla et al., 2018; Fonseca-Aguilar et al., 2015); y se sabe que en Colombia, en particular, existen pocos estudios a nivel psicométrico en adolescentes, ya que, en su mayoría,

las aproximaciones están dirigidas a población clínica o infantil (Mora et al., 2013).

En cuanto a la evaluación de la cognición matemática, algunas escalas usadas en Colombia en el campo clínico e investigativo son el TEDI-MATH (Sueiro & Pereira, 2015), el TEMA-3 (Núñez & Lozano, 2007) y el Pro-cálculo (Feld et al., 2006), sin embargo, estas escalas no cuentan con validación a población colombiana y están diseñadas para niños. Para los adolescentes existen test validados en el contexto colombiano, como los subtest numéricos de la ENI-2 (Matute et al., 2007) o el subtest aritmético de del WISC-IV (Wechsler, 2007), pero estos se concentran solo en evaluar habilidades numéricas básicas.

Por su parte, la subprueba de cálculo matemático del *Wide Range Achievement Test-4* (WRAT-4, Wilkinson & Robertson, 2006) evalúa distintos componentes de la cognición matemática con ejercicios de fracciones, decimales y problemas algebraicos, los cuales responden a una evaluación avanzada de la cognición matemática en edades adolescentes. Sin embargo, esta prueba fue diseñada originalmente en Estados Unidos con una muestra de 3021 participantes (de entre 5 y 94 años de edad).

Es importante tener en cuenta que este test tiene dos formas, y que la *forma B* (azul) es la que evalúa las habilidades matemáticas y tiene niveles adecuados de confiabilidad y validez, con una consistencia interna de .90 a .91 y una validez concurrente que oscila entre .50 y .96 (Wilkinson & Robertson, 2006). De hecho, la *forma B* del WRAT-4 fue validada y estandarizada al español en población mexicana —con 22 de los 40 reactivos originales— por Abreu-Mendoza et al. (2018), quienes evaluaron a 1318 estudiantes de primer año de preparatoria y encontraron, por medio de métodos de validez y confiabilidad de la teoría clásica de los test y la teoría de respuesta al ítem, propiedades psicométricas adecuadas, así como una estructura de tres factores (álgebra básica, números racionales y aritmética).

Ahora bien, no basta con aplicar estos métodos psicométricos, sino que también resulta necesario obtener datos normalizados de las pruebas de evaluación psicológica en el contexto local por medio de la *baremación* —una opción menos costosa y dispendiosa que la creación de test (Aragón, 2012)—, con el fin de obtener resultados específicos de un individuo y poder compararlos con las puntuaciones de un grupo normativo que tenga sus mismas características (Argibay, 2006).

En conclusión, a pesar de la importancia de las habilidades matemáticas para el logro académico, en

el contexto colombiano son pocas las pruebas que las evalúan, y las que se utilizan no tienen puntuaciones estandarizadas, lo cual es esencial para identificar problemas en adolescentes y contar con diagnósticos fiables. En este contexto, en la presente investigación se planteó el objetivo de determinar las propiedades psicométricas de la *forma B* de la subprueba de matemáticas del

WRAT-4 (Wilkinson & Robertson, 2006) en adolescentes colombianos y obtener los baremos de la subprueba en esta población. Para ello, se propuso realizar un análisis de ítems en términos de discriminación y dificultad, evaluar la validez y confiabilidad de la prueba, y obtener puntuaciones estandarizadas según las características sociodemográficas de los grupos estudiados.

Método

Tipo de estudio

Se diseñó un estudio de tipo no experimental, transversal, cuantitativo, descriptivo, y de corte psicométrico, debido a que se buscó especificar las propiedades psicométricas y las características de un instrumento de medición. Específicamente, se obtuvieron los baremos para una muestra en particular (American Educational Research Association [AERA], 2018).

Participantes

La muestra final estuvo conformada por 722 estudiantes de educación básica, de los cuales 390 (54 %) eran de sexo femenino y 332 (46 %) de sexo masculino. La edad promedio de la muestra fue de 15.9 años ($DE = 0.79$), y la distribución de los municipios mostró que 63 participantes (8.7 %) eran de Chameza, 43 (6 %) de Nunchía, 59 (8.2 %) de Sabanalarga, 100 (13.9 %) de Támara, 79 (10.9 %) de Trinidad, y 378 (52.4 %) de la ciudad de Yopal, del departamento de Casanare, Colombia.

La distribución de la muestra por colegios correspondió a la distribución de porcentajes de estudiantes de colegios públicos y privados en Colombia (MinEducación, 2022), por lo cual el 66.8 % (483) asistía a colegios públicos, y el 33.2 % (239) a colegios privados. En particular, el 23.7 % (171) era del grado noveno, 40 % (289) del grado décimo, y 36.3 % (262) del grado once.

Como criterios de inclusión se tuvieron en cuenta: (a) ser estudiantes de un colegio público o privado del departamento de Casanare; (b) tener edades entre los 15 y los 17 años; (c) tener autorización de sus padres o del representante legal para participar en el estudio —por medio de la firma del consentimiento informado—; y (d) no presentar déficit cognitivo, alteraciones perceptivas o visuales, ni déficits auditivos o motores que les impidan realizar la prueba —que se constató con el diligenciamiento de una historia clínica—.

Instrumentos

Formato de historia clínica ad hoc

Se utilizó una historia clínica que diligenciaron los padres de familia, que permitió obtener los datos sociodemográficos de los participantes. Con esta información se constató el cumplimiento de los criterios de inclusión y se seleccionó la muestra que haría parte del estudio.

Cálculo matemático del WRAT-4 (forma B)

Se utilizó la *forma B* de cálculo matemático del *Wide Range Achievement Test-4 (WRAT-4)*, de Wilkinson y Robertson (2006), que evalúa el desempeño matemático de los adolescentes y está compuesta por 40 reactivos que pueden ser administrados de forma individual o grupal, en los cuales los participantes tienen que resolver problemas matemáticos con dificultad incremental en un límite de 15 minutos.

Para este estudio psicométrico se empleó la versión en español de la prueba, adaptada para México por Abreu-Mendoza et al. (2018), en la cual se hizo la traducción de las instrucciones y la modificación de tres ítems en cuanto a su notación matemática, específicamente en los símbolos de división y fracciones. La subprueba inicia con una adición simple y termina con la reducción de una operación racional a su menor expresión.

Esta prueba, en su versión original, cuenta con un índice de consistencia interna general de $\alpha = .85$, y con validez de constructo adecuada a partir de un modelo de tres factores: *aritmética*, *fracciones* y *álgebra*, y *números racionales* (Wilkinson & Robertson, 2006). La validación mexicana obtuvo índices de consistencia interna que oscilaron entre aceptables y buenos ($\alpha \geq .70$), adecuada validez por medio del análisis factorial confirmatorio ($CFI = .97$, $TLI = .96$, $RMSEA = .02$), validez de criterio ($r = .52$, $p < .001$), y validación por el método de Rasch ($infit \geq 0.75$) (Abreu-Mendoza et al., 2018).

Análisis estadísticos

En un primer momento, se realizaron análisis de ítems usando los valores de *discriminación* y dificultad de la prueba. Para calcular la discriminación se utilizó el índice de discriminación, la correlación biserial y la correlación biserial de punto corregida, en donde, en cuanto a los índices y coeficientes, las puntuaciones $< .01$ indicaron valores de discriminación inadecuados; puntuaciones entre $.01$ y $.19$ indicaron valores de discriminación pobres; puntuaciones entre $.20$ y $.29$ indicaron niveles de discriminación regulares; puntuaciones entre $.30$ a $.38$ indicaron valores adecuados de discriminación; y puntuaciones $> .39$ indicaron valores adecuados del ítem. Se eliminaron los ítems que tuvieron puntuaciones pobres e inadecuadas en dos valores de discriminación (Cipresso & Immekus, 2022). Por otra parte, para el análisis de *dificultad* de los ítems se tomaron los valores de: ítem muy fácil ($> .75$), fácil (entre $.55$ y $.74$), normal (entre $.45$ y $.54$), difícil (entre $.20$ y $.44$), y muy difícil ($< .20$) (Cipresso & Immekus, 2022).

Posteriormente, se evaluó la validez de constructo del instrumento por medio del *análisis factorial exploratorio (AFE)* y el *análisis factorial confirmatorio (AFC)*. Para esto, se dividió aleatoriamente la muestra en dos grupos, según las recomendaciones de Lloret-Segura et al. (2014), y con el primer grupo ($n = 361$) se realizó el AFE, mientras que, con el segundo ($n = 361$), el AFC. Para cada submuestra se procuró tener distribuciones homogéneas en cuanto al sexo, edad y grado académico.

En cuanto al AFE, se usaron estadígrafos preliminares de implementación como la prueba de esfericidad de Bartlett ($p < .05$), que ayudó a determinar la existencia de factores dentro del modelo. Para establecer si la muestra era apta para realizar el análisis factorial, se usó la medida de adecuación de muestreo Kaiser-Meyer-Olkin (KMO), en la cual se tomaron los siguientes valores: inaceptable ($0-.50$), mediocre ($.51-.70$), aceptable ($.71-.80$), bueno ($.81-.90$), y excelente ($> .91$) (Fabrigar & Wegener, 2011). Después de esto, se evaluaron las comunalidades y los factores obtenidos en la prueba mediante la proporción de varianza explicada y los autovalores de las cargas por medio del método de extracción de componentes principales. Asimismo, se realizó un método de rotación oblicua (oblimin/promax), asumiendo que las variables latentes estaban correlacionadas. Finalmente, se tomó una medida de corte de las cargas factoriales $> .5$ para asumir que el reactivo cargó en un factor determinado (Dugard et al., 2022).

A continuación se realizó el AFC para probar el modelo obtenido en el AFE, y, asimismo, evaluar los modelos obtenidos en la literatura: el de la validación original y el del estudio realizado en México. Para esto, se obtuvieron los valores de CMIN, en donde se esperó un valor $p \geq .05$. Como dichos valores pueden ser sensibles al tamaño de la muestra, se utilizaron, a su vez, valores de adecuación del modelo CFI, TFI y NFI, donde se esperó obtener valores $> .9$ para indicar un buen ajuste del modelo. De igual forma, se obtuvieron valores de error de la aproximación RMSEA, donde se esperó un valor aproximado a cero, con un punto de corte mínimo de $.06$. Para comparar los modelos obtenidos en los diferentes estudios se utilizó el estadístico de Aiken (AIC), en donde el modelo que presentara un menor valor indicaba un mejor ajuste factorial (Hoyle, 2012).

Después de esto, se realizaron análisis de *confiabilidad* por medio del K-20 para variables numéricas discretas, así como el omega de McDonald, que permite tener una medida más adecuada de la confiabilidad basada en cargas factoriales (Hayes & Coutts, 2020). Valores de K2-20 y $\omega < .7$ indicaron una confiabilidad baja, mientras que valores de KR-20 y $\omega |.70-.89|$ indicaron nivel de confiabilidad adecuada, y valores de KR-20 y $\omega > .90$ indicaron una consistencia interna alta.

Por último, para obtener los *datos normativos* de la prueba se evaluó el efecto de las variables de género y edad por medio de un análisis de varianza (ANOVA factorial de 3×2), para con esto obtener los baremos por medio del cálculo de percentiles.

Para los análisis de confiabilidad, validez convergente, AFE y la obtención de los baremos se utilizó el *software* estadístico SPSS (v.23), y para el AFC se utilizó la extensión AMOS del SPSS (v.23).

Procedimiento

Inicialmente, se contactó a distintas instituciones educativas de seis municipios del departamento de Casanare, Colombia, en las cuales se invitó a participar a los estudiantes mediante la firma del consentimiento informado por parte de los padres de familia, junto con el diligenciamiento del formato de historia clínica. La aplicación de la prueba fue realizada por estudiantes de décimo semestre del programa de psicología de la Universidad Autónoma de Bucaramanga, extensión Unisangil (Yopal), pertenecientes al semillero de investigación de Neuropsicología Básica y Aplicada (Neurobap), así como por el director del semillero, quien también se ocupó de la aplicación,

capacitación a los semilleros en la aplicación de la prueba, la supervisión de las aplicaciones y el manejo de las bases de datos. Las aplicaciones fueron realizadas en horas de la mañana en las aulas escolares de los estudiantes de manera grupal, con una duración aproximada de 25 minutos por grupo (15 minutos para la aplicación total del test). Se controló el efecto del ruido y la iluminación de las aulas.

Aspectos éticos

Este proyecto cumplió con los estándares éticos de investigación en psicología en Colombia (Ley 1090 de 2006), por lo cual se tuvo en cuenta el uso de técnicas válidas, respeto por la dignidad, confidencialidad, participación voluntaria y consentimiento informado. Según la Resolución 8430 de 1993, la investigación fue considerada como de riesgo mínimo.

Resultados

En la Tabla 1 se presentan los análisis de los reactivos en cuanto al índice y coeficientes de discriminación, así como el análisis de dificultad de los ítems. A partir de las puntuaciones, los ítems 6, 33, 35, 36, 38 y 40 fueron eliminados de la escala por tener valores pobres e inadecuados. En cuanto a la dificultad, los

valores muestran que los ítems 1-12, 14 y 15 fueron considerados como fáciles, mientras que los ítems 13, 16, 18, 21-23, 25 y 26 fueron categorizados como de dificultad normal, los ítems 17, 19, 24 y 28 se consideraron difíciles, y los ítems 27, 29-32, 34, 37 y 39 fueron catalogados como muy difíciles.

Tabla 1. Análisis de ítems de la subprueba de matemáticas del WRAT-4

Reactivo	Índice de discriminación	Correlación biserial	Correlación biserial corregida	Índice de dificultad
R_1	.05	.29	.26	.98
R_2	.10	.28	.24	.96
R_3	.07	.34	.31	.98
R_4	.05	.31	.28	.98
R_5	.08	.31	.27	.97
R_6	.09	.23	.18	.95
R_7	.14	.31	.25	.92
R_8	.16	.29	.24	.93
R_9	.20	.30	.24	.88
R_10	.11	.37	.32	.97
R_11	.13	.27	.22	.94
R_12	.22	.42	.38	.93
R_13	.52	.47	.40	.75
R_14	.38	.37	.27	.76
R_15	.38	.39	.32	.81
R_16	.52	.44	.35	.66
R_17	.34	.36	.28	.20

Reactivo	Índice de discriminación	Correlación biserial	Correlación biserial corregida	Índice de dificultad
R_18	.63	.52	.43	.53
R_19	.44	.47	.39	.22
R_20	.13	.26	.21	.06
R_21	.52	.45	.37	.66
R_22	.59	.52	.44	.34
R_23	.62	.53	.45	.36
R_24	.34	.37	.29	.22
R_25	.57	.41	.31	.42
R_26	.53	.49	.41	.29
R_27	.10	.31	.27	.04
R_28	.34	.41	.34	.20
R_29	.40	.54	.48	.15
R_30	.30	.49	.44	.10
R_31	.26	.30	.23	.15
R_32	.17	.40	.36	.07
R_33	.17	.26	.19	.10
R_34	.20	.36	.31	.09
R_35	.01	.12	.11	.01
R_36	.11	.22	.17	.07
R_37	.16	.38	.34	.06
R_38	.00	.11	.11	.00
R_39	.08	.28	.24	.03
R_40	.02	.18	.17	.01

En cuanto a la validez de constructo del test, se realizó un análisis factorial exploratorio con los 34 reactivos resultado del análisis de ítems llevado a cabo con la primera submuestra ($n = 361$). Los hallazgos mostraron la existencia de tres factores que explicaron el 32.97 % de la varianza (Bartlett = 4768.83 $p < .001$). Asimismo, la muestra se ajustó a los resultados del análisis (KMO = .88).

Las cargas factoriales con el método de rotación oblicua de cada uno de los reactivos se pueden ob-

servar en la Tabla 2. Específicamente, el primer factor fue conformado por los ítems 1-5 y 7-11, que evalúan el constructo de *aritmética básica*; el segundo factor lo conforman los ítems 19, 27-29, 30, 32, 34, 37 y 39, que evalúan el constructo de *números racionales*; y el tercer factor incluye los ítems 12-16, 18, 21-23, 25, 26, que evalúan problemas de *fracciones y ecuaciones algebraicas simples*. Los ítems 20 y 24 no cargaron en ningún factor.

Tabla 2. Cargas factoriales de los ítems en cada uno de los componentes de la subprueba de matemáticas del WRAT-4

Reactivo	Componente		
	1	2	3
R4	.81		
R3	.77		
R1	.74		
R5	.67		
R10	.64		
R2	.59		
R11	.51		
R7	.48		
R8	.44		
R9	.36		
R37		.66	
R32		.66	
R29		.65	
R34		.56	
R30		.55	
R39		.53	
R27		.50	
R28		.49	
R19		.47	
R20	-	-	-
R16			.63
R13			.59
R18			.48
R21			.48
R26			.47
R25			.47
R23			.44
R12			.43
R15			.41
R22			.40
R14			.40
R24	-	-	-

El análisis factorial confirmatorio, que se realizó con la segunda submuestra ($n = 361$) para establecer la validez de constructo de la prueba, se llevó a cabo a partir de tres modelos: uno que incluyó los resultados del modelo factorial exploratorio obtenido, otro que incluyó los ítems de la validación mexicana, y, el tercero, que incluyó los ítems de la validación original de la prueba. Los resultados de los índices de ajuste se pueden observar en la Tabla 3.

A nivel general, el modelo de tres factores con población colombiana tuvo un ajuste adecuado (CMIN/DF < .30) en cada uno de los índices de ajuste (GFI, CFI

y TLI > .90) y un valor adecuado en los valores del error de aproximación (RMSEA < .06).

Por su parte, el modelo que incluyó los ítems de la validación mexicana y el de la prueba original no evidenciaron un ajuste adecuado. Además, el índice de comparación de Aiken (AIC) mostró que el modelo de tres factores colombianos fue el que mejor ajuste presentó, ya que tuvo el menor valor entre los modelos evaluados. La distribución de cada uno de los reactivos en los diferentes factores y los valores de las estimaciones estandarizadas entre los factores se pueden observar en la Figura 1.

Tabla 3. Comparación de ajuste factorial de los modelos de validez de constructo de la subprueba de matemáticas del WRAT-4

Modelos factoriales	CMIN/DF	GFI	CFI	TLI	RMSEA	AIC
Tres factores basados en AFE	1.835	.934	.921	.914	.034	774.384
Tres factores (Abreu-Mendoza et al., 2018)	3.303	.910	.777	.750	.057	864.233
Tres factores (Wilkinson & Robertson, 2006)	3.014	.846	.722	.703	.053	2379.937

Nota. CMIN/DF: razón chi cuadrado sobre grados de libertad. GFI: índice de bondad de ajuste. CFI: índice de ajuste comparativo. TLI: Índice Tucker-Lewis. RMSEA: error cuadrado de aproximación de las raíces. AIC: criterio de comparación de Aiken.

Por otra parte, los análisis de confiabilidad, obtenidos de los análisis tanto para el total de la prueba como para cada uno de los componentes, mostraron valores adecuados de consistencia in-

terna. En general, los valores del total de la prueba tuvieron una consistencia interna buena, y cada uno de los componentes presentó consistencia interna aceptable (véase Tabla 4).

Figura 1. Modelo factorial de la subprueba de matemáticas del WRAT-4 para cada uno de los ítems

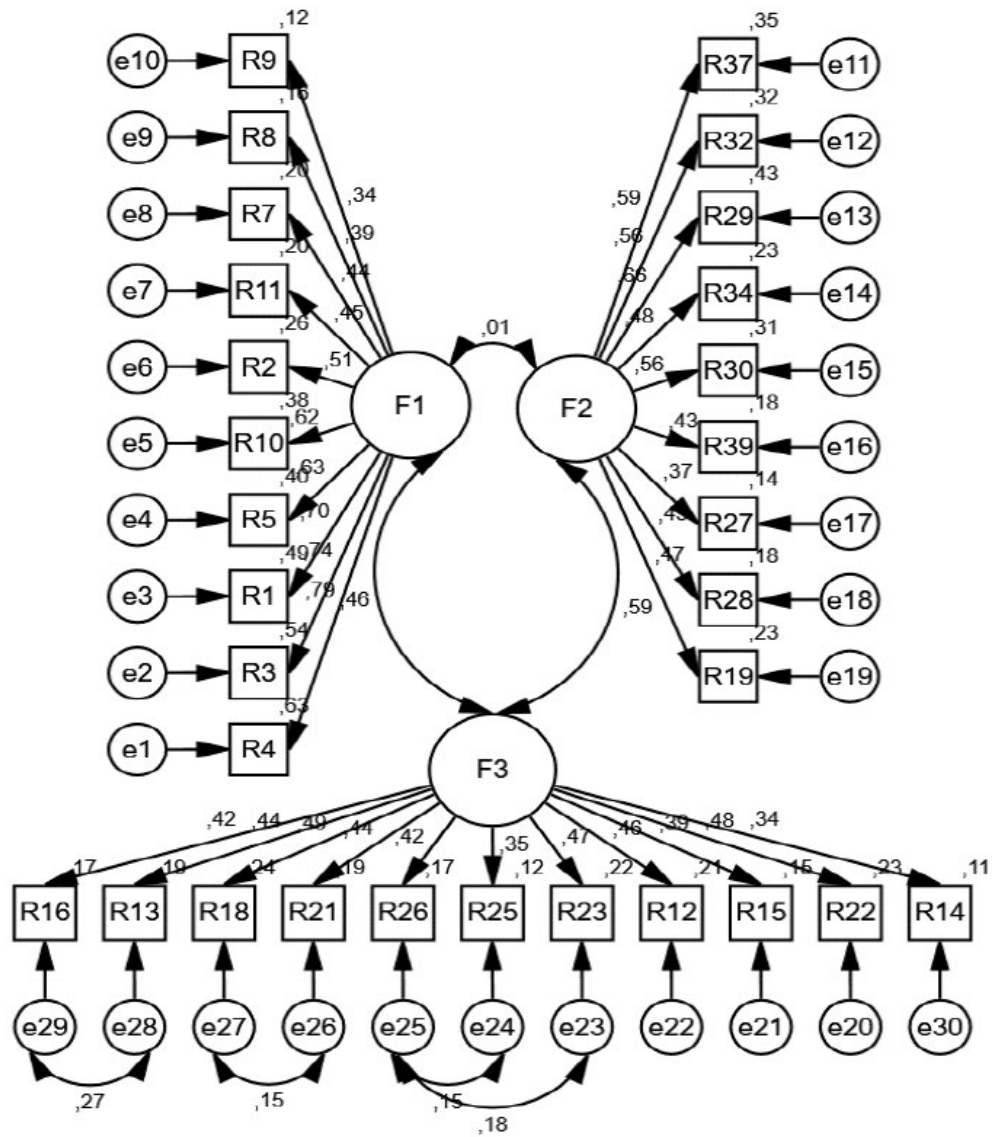


Tabla 4. Análisis de confiabilidad de la subprueba de matemáticas del WRAT-4 total y por factores

Factor	Ítems	KR-20	ω
Aritmética básica	1-5, 7-11.	.76	.77
Números racionales	19, 27-30, 32, 34, 37 y 39.	.74	.75
Problemas con fracciones y álgebra	12-16, 18, 21, 23, 25 y 26.	.72	.70
WRAT Total		.81	.71

Teniendo en cuenta los resultados de la validez y confiabilidad, en donde se observó que el WRAT-4 es una prueba es válida y confiable, se realizaron análisis de efectos para obtener los baremos de la prueba. Por tanto, se evaluó el efecto de las variables de sexo, grado académico y tipo de colegio sobre el resultado total de la prueba, con lo cual se encontró que no hubo diferencias estadísticamente significativas entre hombres ($M = 16.68, DE = 4.17$) y mujeres ($M = 16.61, DE = 3.49$) en la puntuación total de la prueba ($t(720) = -0.26, p = .399, d$ de Cohen = 0.02); y que el tipo de colegio mostró diferencias estadísticamente significativas en cuanto a la puntuación total de la prueba ($t(720) = -11.14, p < .001, d$ de Cohen = 1.26), en donde los participantes del colegio

privado ($M = 20.54, DE = 3.89$) tuvieron puntuaciones promedio mayores que los participantes de colegio público ($M = 16.09, DE = 3.47$).

Finalmente, se realizó un ANOVA factorial para identificar si existen diferencias entre el colegio y el grado académico sobre las puntuaciones del WRAT-4, y efectivamente se encontraron diferencias ($F(2) = 13.09, p < .001, \eta^2 = .04, \beta-1 = .99$), pues, a medida que aumentó el grado académico, aumentaron las puntuaciones del total de la prueba (véase Tabla 5). Por último, teniendo en cuenta los efectos de las variables, se establecieron los puntos de corte percentiles para cada tipo de colegio y grado académico (véase Tabla 6).

Tabla 5. Efectos del tipo de colegio y grado académico sobre las puntuaciones de la subprueba de matemáticas del WRAT-4 total

Grado	Tipo de colegio	M	DE	Post hoc
Público	Noveno (9°)	14.58	3.96	
	Décimo (10°)	15.84	3.33	9° < 10° < 11°
	Once (11°)	16.82	3.47	
Privado	Noveno (9°)	18.67	6.50	
	Décimo (10°)	19.41	3.52	9° < 10° < 11°
	Once (11°)	22.34	3.59	

Nota. Post hoc colegio público 9° vs. 10°: $p = .046$, [IC 95 % = -2.51, -0.14]; 9° vs. 11°: $p < .001$, [IC 95 % = -3.53, -0.95]; 10° vs. 11°: $p = .003$, [IC 95 % = -1.67, -0.27]. Post hoc colegio privado 9° vs. 10°: $p = .003$, [IC 95 % = -1.75, -0.13]; 9° vs. 11°: $p < .001$, [IC 95 % = -3.68, -0.02]; 10° vs. 11°: $p < .001$, [IC 95 % = -4.74, -1.12].

Tabla 6. Cálculo de los percentiles de los tipos de colegio (público/privado) y grados académicos

Puntuación percentil	Puntuación directa					
	Colegio público			Colegio privado		
	Noveno	Décimo	Once	Noveno	Décimo	Once
1	-	0-2	0-1	-	-	-
2	-	3-9	2-9	-	-	-
3	0-2	10	10-11	-	0-11	-
4	3-4	-	12	-	-	-
5	5-7	-	-	-	12	0-16

Puntuación percentil	Puntuación directa					
	Colegio público			Colegio privado		
	Noveno	Décimo	Once	Noveno	Décimo	Once
6	-	11	-	-	-	-
7	-	-	-	-	13	-
8	-	-	-	-	-	17
11	-	12	13	-	14	-
13	10	-	-	-	15	-
16	-	-	-	-	-	18
17	11	-	-	-	-	-
21	12	13	14	-	16	-
24	-	-	-	-	17	-
28	-	-	-	0-12	-	-
30	-	-	-	-	-	19
31	-	-	15	-	-	-
32	-	-	-	13	-	-
33	-	14	-	-	-	-
35	-	-	-	14	-	-
37	13	-	-	-	-	-
39	-	-	-	15	-	-
40	-	-	-	-	18	-
41	-	-	-	-	-	20
42	-	-	-	16	-	-
43	14	-	-	-	-	-
44	-	-	-	-	-	21
45	-	15	-	-	-	-
46	-	-	16	17	-	-
49	-	-	-	18	-	-
53	-	-	-	-	19	-
54	-	-	-	19	-	-
55	-	-	-	-	-	22
56	-	16	-	-	-	-
58	-	-	-	20	-	-
61	-	-	17	-	-	-
62	15	-	-	21	-	-

Puntuación percentil	Puntuación directa					
	Colegio público			Colegio privado		
	Noveno	Décimo	Once	Noveno	Décimo	Once
63	-	-	-	-	-	23
65	-	-	-	-	20	-
66	-	-	-	22	-	-
69	-	17	-	-	-	24
70	16	-	-	23	-	-
72	-	-	18	-	-	-
73	-	-	-	-	21	-
74	-	-	-	≥ 24	-	-
77	-	-	-	-	-	25
79	-	18	-	-	-	-
80	-	-	19	-	-	26
82	-	-	-	-	22	-
84	17	-	-	-	23	-
86	18	-	-	-	-	-
88	-	-	-	-	24	27
89	-	19	20	-	-	-
92	19	-	21	-	-	-
93	-	20	-	-	-	-
94	20	-	-	-	-	-
95	-	-	22	-	≥ 25	-
96	21	21	-	-	-	-
97	-	-	23	-	-	≥ 28
98	≥ 22	22	-	-	-	-
99	-	≥ 23	≥ 24	-	-	-

Discusión

El objetivo de este trabajo fue determinar las propiedades psicométricas y obtener los baremos de la subprueba de matemáticas del *Wide Range Achievement Test-4* (WRAT-4, Wilkinson & Robertson, 2006), en adolescentes colombianos. En general, se encontró que el subtest es un instrumento válido y confiable para evaluar las habilidades matemáticas en estudiantes de 15 a 17 años en el contexto colombiano, y, por tanto, se presentan los datos normativos (baremos), que pueden ser utilizados para la evaluación de adolescentes con el fin de identificar problemas matemáticos.

En particular, respecto al análisis de ítems, en el presente estudio, al igual que en el de Abreu-Mendoza et al. (2018), se identificó que algunos de los reactivos de la subprueba no presentaron niveles de discriminación, dificultad o carga factorial aceptables, y por tanto tuvieron que ser eliminados. Dicha eliminación permite tener una prueba capaz de discriminar si los individuos que presentan un rasgo o constructo medido también responden al ítem de manera similar (Gyamfi & Wren, 2022), por lo que se puede afirmar que la subprueba tiene la capacidad de identificar estudiantes con habilidades y con problemas matemáticos. Así mismo, se identificó que la subprueba tuvo una distribución proporcional entre ítems fáciles, medios y difíciles, lo que garantiza la evaluación adecuada del constructo que se pretende evaluar (Hurtado, 2018).

Ahora bien, es de señalar que la muestra colombiana no respondió acertadamente a algunos de los reactivos que evaluaron fracciones y álgebra, similar a lo encontrado en la validación mexicana (Abreu-Mendoza et al., 2018), pero se debe tener en cuenta que Colombia se ubica por debajo del promedio en el rendimiento matemático según los resultados de distintas pruebas que buscan evaluar el rendimiento matemático en distintos países, como PISA, SERCE y TIMSS (MinEducación, 2021; OCDE, 2018). Por esta razón, dichos reactivos de la subprueba fueron catalogados como *reactivos difíciles* en el análisis de ítems, lo cual puede ser explicado por los bajos niveles de desempeño matemático en la muestra, que a su vez podrían estar explicados por contingencias inherentes de la adquisición de los números racionales y la aritmética, y contingencias culturales de la forma en que son enseñadas las matemáticas en el país (Siegler & Lortie-Forgues, 2017).

Teniendo en cuenta este factor cultural, el análisis de la validez de constructo arrojó la existencia de tres

factores: *aritmética, números racionales, y problemas aritméticos, fracciones y ecuaciones simples*, de manera similar a lo encontrado en el estudio de validación de México (Abreu-Mendoza et al., 2018). En la prueba original, aunque no se realizaron modelos de ecuaciones estructurales, los factores asumidos por los investigadores tenían una distribución de ítems diferente a la encontrada en el presente estudio, ya que el modelo factorial no tuvo un ajuste adecuado para la muestra colombiana. De este modo, se podría inferir un efecto cultural entre las validaciones de Latinoamérica, sobre todo si se tiene en cuenta la dificultad en la resolución de problemas aritméticos por parte de los estudiantes de ambos países, reflejada en los resultados de pruebas internacionales como las pruebas PISA (OCDE, 2018).

Por otra parte, se debe tener en cuenta que la carga factorial de los ítems en tres factores distintos puede obedecer a la naturaleza de los reactivos que se evalúan en la prueba. En este sentido, se ha visto que la ejecución de tareas y problemas aritméticos complejos requieren más recursos de la memoria de trabajo que la resolución de problemas simples, así como la activación de regiones cerebrales distintas, teniendo en cuenta el tipo de problema matemático que se lleva a cabo (Arsalidou & Taylor, 2011). Por ejemplo, Artemenko et al. (2018) encontraron la activación de la red de la corteza parietal bilateral al momento de realizar operaciones básicas de sumas y multiplicaciones, y que se requiere de la activación de otras regiones, como la corteza prefrontal y la circunvolución fusiforme, para la resolución de problemas complejos aritméticos. Otras estructuras como el surco intraparietal bilateral y la circunvolución angular izquierda se han identificado como soporte en las primeras etapas del procesamiento matemático según estudios de neuroimagen con señales BOLD (González & Hornauer-Hughes, 2014; Looi et al., 2016).

De igual forma, el modelo factorial obtenido sugiere que las habilidades matemáticas de los adolescentes no son un dominio global, por lo que el entendimiento y procesamiento de los distintos tipos de problemas aritméticos, ya sean para resolver fracciones o problemas de álgebra básica, varían y también dependen de los factores de aprendizaje y adquisición de las habilidades matemáticas (Siegler & Lortie-Forgues, 2017).

Así mismo, se pudo evidenciar que la consistencia interna de los resultados de las puntuaciones en la

subprueba del WRAT-4 total, así como en cada uno de sus factores, fueron adecuados, por lo que, además de concluir que la escala es confiable y útil para la aplicación en población colombiana, los índices de confiabilidad indican que hubo respuestas consistentes en los ítems de la prueba (Hayes & Coutts, 2020). Esto daría a entender que, en su mayoría, los adolescentes colombianos tienen un nivel similar de rendimiento matemático. Los resultados de consistencia interna del test también fueron adecuados en las otras validaciones (Abreu-Mendoza et al., 2018; Wilkinson & Robertson, 2006).

Finalmente, se obtuvieron las puntuaciones normativas percentiles, divididas según el tipo de colegio y el grado académico. Dichas diferencias entre el desempeño académico y matemático entre colegios públicos y privados ya se han reportado en estudios anteriores (Rosselli & Matute, 2011; Zacharopoulos et al., 2021) y son explicadas por factores sociales, como el acceso a herramientas, una calidad educativa mejor, y a la escolaridad de los padres (Silver & Libertus, 2022). Asimismo, las diferencias encontradas entre los grados académicos se explican por el efecto que tienen las diferencias curriculares entre cada uno de los años de escolaridad tenidos en cuenta (Siegler & Lortie-Forgues, 2017), lo que generaría cambios dinámicos de plasticidad cerebral en

estructuras como el lóbulo parietal, temporal, prefrontal y el bucle córtico-talámico, resultado de la estimulación y el aprendizaje de los diferentes temas matemáticos que se ven en los distintos grados académicos (Arsalidou & Taylor, 2011; Gilmore et al., 2018).

Contar con puntuaciones normativas divididas según las diferencias sociodemográficas de la muestra normativa permite disponer de una prueba que puede ser realizada para grupos de estudiantes con características específicas, discriminar entre estudiantes con posibles problemas matemáticos, así como reducir los sesgos de interpretación asociados a las diferencias sociales y culturales de cada uno de los grupos evaluados.

En conclusión, se puede afirmar que la *forma B* de la subprueba de matemáticas del WRAT-4 resulta ser una prueba válida y confiable para evaluar distintos componentes de las habilidades matemáticas de los estudiantes colombianos de los últimos grados académicos, y que, a partir de este estudio, se cuenta con baremos para interpretar mejor los resultados obtenidos con la aplicación de la prueba. Para futuros estudios se recomienda complementar la evaluación psicológica con otras pruebas que pongan de manifiesto la existencia de problemas matemáticos, para su correspondiente intervención.

Referencias

- Abreu-Mendoza, R. A., Chamorro, Y., & Matute, E. (2018). Psychometric properties of the WRAT math computation subtest in Mexican adolescents. *Journal of Psychoeducational Assessment*, 37(8), 957-972. <https://doi.org/10.1177/0734282918809793>
- American Educational Research Association [AERA]. (2018). *Standards for educational and psychological testing*. AERA.
- Aragón, L. (2012). *Evaluación Psicológica. Historia, Fundamentos Teóricos, conceptuales y Psicometría* (1.ª ed.). Editorial El Manual Moderno.
- Arango-Lasprilla, J., Rodríguez-Irizarry, W., Oliveras-Rentas, R., & Olabarrieta-Landa, L. (2018). La práctica de la neuropsicología en Puerto Rico: implicaciones para la certificación de la especialidad. *Revista Iberoamericana de Neuropsicología*, 1(1), 45-62. <https://neuropsychologylearning.com/portfolio-item/neuropsicologia-en-puerto-rico/>
- Argibay, J. (2006). Técnicas psicométricas, Cuestiones de validez y confiabilidad. *Subjetividad y Procesos Cognitivos*, (8), 15-33. http://dspace.uces.edu.ar:8180/dspace/bitstream/handle/123456789/765/T%C3%A9cnicas_psicom%C3%A9tricas.pdf?sequence=1
- Arsalidou, M., & Taylor, M. J. (2011). Is $2 + 2 = 4$? Meta-analyses of brain areas needed for numbers and calculations. *Neuroimage*, 54(3), 2382-2393. <https://doi.org/10.1016/j.neuroimage.2010.10.009>
- Artemenko, C., Soltanlou, M., Ehlis, A. C., Nuerk, H. C., & Dresler, T. (2018). The neural correlates of mental arithmetic in adolescents: a longitudinal fNIRS study. *Behavioral and brain functions*, 14, 5. <https://doi.org/10.1186/s12993-018-0137-8>
- Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences*, 14(12), 534-541. <https://doi.org/10.1016/j.tics.2010.09.007>
- Cipresso, P., & Immekus, J. C. (Eds.). (2022). *Statistical Guidelines: New Developments in Statistical Methods and Psychometric Tools*. Frontiers Research Topics.
- Cohen, R., & Swerdlik, M. (2001). *Pruebas y evaluación psicológica* (4.ª ed.). McGraw-Hill.
- Devine, A., Soltész, F., Nobes, A., Goswami, U., & Szücs, D. (2013). Gender differences in developmental dyscalculia depend on diagnostic criteria. *Learning and Instruction*, 27, 31-39. <https://doi.org/10.1016/j.learninstruc.2013.02.004>
- Dorneles, B. V. (2019). Mathematical learning and its difficulties in Latin-American countries. En A. Fritz, V. G. Haase & P. Räsänen (eds.), *International handbook of mathematical learning difficulties* (pp. 201-212). Springer.
- Dugard, P., Todman, J., & Staines, H. (2022). *Approaching multivariate analysis: A practical introduction* (2ª ed.). Taylor & Francis.
- Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.
- Feld, V., Taussik, I., & Azzareto, C. (2006). *Test PRO-CÁLCULO. Test para la evaluación del procesamiento del número y el cálculo en niños*. Paidós.
- Fonseca-Aguilar, P., Olabarrieta-Landa, L., Rivera, D., Aguayo, A., Ortiz, X. A., Rabago, B. V., Agudelo, Y., Álvarez, E., & Arango-Lasprilla, J. C. (2015). Situación actual de la práctica profesional de la neuropsicología en México. *Psicología desde el Caribe*, 32(3), 344-364. <http://www.scielo.org.co/pdf/psdc/v32n3/v32n3a02.pdf>
- Frangou, P., Correia, M., & Kourtzi, Z. (2018). GABA, not BOLD, reveals dissociable learning-dependent plasticity mechanisms in the human brain. *Elife*, 7, e35854. <https://doi.org/10.7554/eLife.35854>

- Gilmore, C., Göbel, S. M., & Inglis, M. (2018). *An Introduction to Mathematical Cognition K* (1.ª ed.). Taylor y Francis Group. <https://doi.org/10.4324/9781315684758>
- González, R., & Hornauer-Hughes, A. (2014). Cerebro y lenguaje. *Revista Hospital Clínico Universidad de Chile*, 25(1), 144-153. https://www.redclinica.cl/Portals/o/Users/014/14/14/Cerebro_%20y_lenguaje.pdf
- Gyamfi, A., & Wren, D. G. (2022). Determining the Difficulty and Discrimination Parameters of a Mathematics Performance-Based Assessment. *Creative Education*, 13(11), 3483-3489. <https://doi.org/10.4236/ce.2022.1311223>
- Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. *But... Communication Methods and Measures*, 14(1), 1-24. <https://doi.org/10.1080/19312458.2020.1718629>
- Hebben, N., & Milberg, W. (2011). *Fundamentos para la evaluación neuropsicológica* (2.ª ed.). Editorial Manual Moderno.
- Herrera, A. N., & León, F. A. (2022). Desarrollo del Instrumento para Evaluar la Calidad Técnica de Pruebas Psicológicas. *Revista Colombiana de Psicología*, 31(2), 65-76. <https://doi.org/10.15446/rcp.v31n2.93018>
- Hoyle, R. H. (Ed.). (2012). *Handbook of structural equation modeling*. Guilford press.
- Hurtado, L. (2018). Relación entre los índices de dificultad y discriminación. *Revista digital de investigación en docencia universitaria*, 12(1), 273-300. <http://dx.doi.org/10.19083/ridu.12.614>
- Klein, E., Moeller, K., Glauche, V., Weiller, C., & Willmes, K. (2013). Processing pathways in mental arithmetic—evidence from probabilistic fiber tracking. *PloS one*, 8(1), e55455. <https://doi.org/10.1371/journal.pone.0055455>
- Kroesbergen, E. H., van Luit, J. E., van Lieshout, E. C., Van Loosbroek, E., & van de Rijt, B. A. (2009). Individual Differences in Early Numeracy: The Role of Executive Functions and Subitizing. *Journal of Psychoeducational Assessment*, 27(3), 226-236. <https://doi.org/10.1177/0734282908330586>
- Le Corre, M., Van de Walle, G., Brannon, E. M., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology*, 52(2), 130-169. <https://doi.org/10.1016/j.cogpsych.2005.07.002>
- Ley 1090 de 2006. Por la cual se reglamenta el ejercicio de la profesión de Psicología, se dicta el Código Deontológico y Bioético y otras disposiciones. Diario Oficial n.º 46383 del 6 de septiembre de 2006. Congreso de Colombia. <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=66205>
- Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *Anales de psicología/annals of psychology*, 30(3), 1151-1169. <https://doi.org/10.6018/analesps.30.3.199361>
- Looi, C., Thompson, J., Krause, B., & Kadosh, R. C. (2016). The neuroscience of mathematical cognition and learning. *OECD Education Working Papers*, 136. <https://doi.org/10.1787/5jlwmn3ntbr7-en>
- Matute, E., Rosselli, M., Ardila, A., & Ostrosky-Solís, F. (2007). *Evaluación neuropsicológica infantil*. Manual Moderno.
- Ministerio de Educación Nacional de Colombia [MinEducación]. (2022). *El Ministerio de Educación Nacional pone a disposición la información estadística de educación superior 2021*. MinEducación. <https://snies.mineducacion.gov.co/portal/401926>
- Ministerio de Educación Nacional de Colombia [MinEducación] (2 de diciembre de 2021). *Colombia: qué y cómo mejorar a partir de la prueba PISA*. MinEducación. <https://www.mineducacion.gov.co/1621/article-162392.html>
- Mora, C. N., Sierra, M. S., & Castro, C. (2013) *Revisión de pruebas psicológicas para niños validadas o estandarizadas en Colombia* [Tesis de grado]. Universidad de La Sabana, Colombia. <https://intellectum.unisabana.edu.co/bitstream/handle/10818/9552/MAR%c3%8dA%20SOLEDAD%20SIERRA%20%20tesis.pdf?sequence=1&isAllowed=y>

- Núñez, M., & Lozano, I. (2007). *The test of early mathematics ability*. TEMA-3. TEA Ediciones.
- Organización para la Cooperación y el Desarrollo Económicos [OCDE]. (2018). PISA: 2018. Insights and interpretations. OCDE. <https://www.oecd.org/content/dam/oecd/en/about/programmes/edu/pisa/publications/national-reports/pisa-2018/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf>
- Olabarrieta-Landa L., Caracuel A., Pérez-García M., Panyavin I., Morlett-Paredes A., & ArangoLasprilla J. C. (2016). The profession of neuropsychology in Spain: Results of a national survey. *The Clinical Neuropsychologist*, 30(8), 1335-1355. <https://doi.org/10.1080/13854046.2016.1183049>
- Reigosa-Crespo, V., Valdés-Sosa, M., Butterworth, B., Estévez, N., Rodríguez, M., Santos, E., Torres, P., Suárez, R., & Lage, A. (2012). Basic Numerical Capacities and Prevalence of Developmental Dyscalculia: The Havana Survey. *Developmental Psychology*, 48(1), 123-135. <https://doi.org/10.1037/a0025356>
- Resolución 8430 de 1993. Por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud. Ministerio de Salud. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DIJ/RESOLUCION-8430-DE-1993.PDF>
- Rosselli, M., & Matute, E. (2011). La neuropsicología del desarrollo típico y atípico de las habilidades numéricas. *Revista neuropsicología, neuropsiquiatría y neurociencias*, 11(1), 123-140. <https://dialnet.unirioja.es/servlet/articulo?codigo=3640864>
- Siegler, R. S., & Lortie-Forgues, H. (2017). Hard lessons: Why rational number arithmetic is so difficult for so many people. *Current Directions in Psychological Science*, 26(4), 346-351. <https://doi.org/10.1177/0963721417700129>
- Silver, A. M., & Libertus, M. E. (2022). Environmental influences on mathematics performance in early childhood. *Nature Reviews Psychology*, 1, 407-418. <https://doi.org/10.1038/s44159-022-00061-z>
- Sueiro, J., & Pereña, J. (2015). *TEDI-MATH, Test para el diagnóstico de las Competencias Básicas en Matemáticas* (2.ª ed.). TEA Ediciones.
- Wilkinson, G. S., & Robertson, G. J. (2006). *WRAT 4 Wide Range Achievement Test*. Psychological Assessment Resources.
- Wechsler, D. (2007). *Escala Wechsler de Inteligencia para Niños-IV*. Manual Moderno.
- Zacharopoulos, G., Sella, F., & Kadosh, R. C. (2021). The impact of a lack of mathematical education on brain development and future attainment. *Proceedings of the National Academy of Sciences*, 118(24), e2013155118. <https://doi.org/10.1073/pnas.2013155118>