

Evidencias métricas sobre la Escala de Razonamiento Científico en Colombia: Un ejemplo práctico sobre la importancia de los procesos de adaptación

Psychometric evidence of the Scientific Reasoning Scale in Colombia: A practical example of the importance of adaptation processes



ARTÍCULO DE INVESTIGACIÓN
Copyright © 2025
by PsicoGente

Recibido: 16-05-24
Aceptado: 25-08-25
Publicado: 01-07-25

Tomás Martínez-Rodríguez
Universidad de La Sabana, Chia, Colombia.

María-Fernanda Rincón-Silva
Universidad de La Sabana, Chia, Colombia.

Natalia Tovar-Peña
Universidad de La Sabana, Chia, Colombia

***Fernando Riveros Munévar**
Universidad de La Sabana, Chia, Colombia.
E-mail: edgar.riveros@unisabana.edu.co

Tomás Martínez-Rodríguez  - María-Fernanda Rincón-Silva  -
Natalia Tovar-Peña  - Fernando Riveros Munévar* 

Resumen:

Objetivo: La adaptación de pruebas y el análisis de las propiedades métricas de los instrumentos garantiza la objetividad y justicia en el trabajo profesional del psicólogo, lo que plantea la necesidad de probar las escalas, especialmente, porque algunas de ellas no logran ajustarse a todo contexto. Este estudio instrumental consistió en el proceso de adaptación de la Escala de Razonamiento Científico, en muestras colombianas.

Método: Se llevó a cabo un estudio de tipo instrumental sobre la validación de la Escala de Razonamiento Científico (SRS) de Drummond y Fischhoff (2017). La muestra fue de 264 participantes (muestreo no probabilístico por bola de nieve, 59,84 % mujeres, M de edad de 30,81 años, 55,3 % con nivel socioeconómico medio o alto, sin diagnósticos médicos o psiquiátricos) y un bootstrapping de 1000. Se llevó a cabo un proceso de traducción-contra traducción, validación por jueces, análisis factoriales exploratorios, análisis factoriales confirmatorios, de fiabilidad y de respuesta al ítem.

Resultados: Pese al procedimiento de adaptación aplicado, la escala no se ajustó en modelos factoriales, presentó bajos coeficientes de confiabilidad y el funcionamiento diferencial exhibió bajo nivel de ajuste y separación.

Conclusión: Es indispensable comprobar la fiabilidad de las pruebas psicológicas antes de ser usadas, comprendiendo las diferencias culturales, de idioma, usos del lenguaje y características sociales que puedan influir; esto en pro de garantizar mediciones objetivas y de calidad. Igualmente, se discuten otras posibilidades analíticas que pudieran evaluarse en este tipo de casos.

Palabras clave: adaptación, validez, confiabilidad, psicometría, razonamiento científico, análisis factorial exploratorio, análisis factorial confirmatorio.

Abstract

Objective: The adaptation of psychological assessments and the analysis of their psychometric properties are both essential for ensuring objectivity and fairness in professional psychological practice. This underscores the need to evaluate measurement scales, given that some may not be suitable for all contexts. This instrumental study involved the adaptation of the Scientific Reasoning Scale for use with Colombian samples.

Method: An instrumental study was conducted to validate the Scientific Reasoning Scale (SRS) by Drummond and Fischhoff (2017). The sample consisted of 264 participants (non-probability snowball sampling, 59,84 % women, mean age 30,81 years, 55,3 % from middle or high socioeconomic background, without medical or psychiatric diagnoses) and a bootstrapping of 1000. A translation-counter-translation process, expert validation, exploratory factor analysis, confirmatory factor analysis, reliability analysis, and item response analysis were performed.

Results: Despite the rigorous adaptation procedures, the scale did not exhibit a good fit in the factor models, showed low reliability coefficients, and displayed poor fit and discrimination in the differential item functioning analysis.

Conclusions: It is essential to verify the reliability of psychological tests before they are used, considering cultural differences, language, usage patterns, and social characteristics that may influence the results; this is to ensure objective and high-quality measurements. Other analytical approaches that could be considered in these types of cases are also discussed.

Keywords: adaptation, validity, reliability, psychometrics, scientific reasoning, exploratory factor analysis, confirmatory factor analysis

Cómo citar este artículo (APA):

Martínez-Rodríguez, T., Rincón-Silva, M.F., Tovar-Peña, N. & Riveros Munévar, F. 2025. Evidencias métricas sobre la escala de razonamiento científico en Colombia: un ejemplo práctico sobre la importancia de los procesos de adaptación. *PsicoGente* 28(54), 1-12. <https://doi.org/10.17081/psico.28.54.7253>

1. INTRODUCCIÓN

Las pruebas psicométricas son esenciales para explicar científicamente la conducta humana, puesto que son el medio que permite observarla y cuantificarla objetivamente (Livia & Ortiz, 2014). En cuanto a las variables que pueden medir este tipo de pruebas, debido al carácter psicológico de las mismas, se habla de conductas observables y representativas de cierto fenómeno dentro de la psicología, las cuales pueden verse reflejadas en una acción (Martínez & Villota, 2022). Para que estas puedan ser cuantificadas, deben de ser manifestadas por un individuo mediante atributos, características, habilidades y/o rasgos que se miden por medio de la expresión conductual externa y observable, usando las pruebas psicométricas (Martínez & Villota, 2022).

Al mismo tiempo, se comprende que las conductas esperadas de las personas componen el constructo a medir (Sarabia & Alconero, 2019), que se vuelven tangibles al ser medidas por ítems o reactivos –preguntas o afirmaciones que miden rasgos psicológicos– manifestados por los individuos (Morucci *et al.*, 2024). Al respecto, un ítem es una representación conductual de dicho conjunto de rasgos que miden una conducta particular, por ende, las conductas medidas deben estar reflejadas en el ítem, de forma tal que el ítem sea comprensible y representativo del constructo que se pretende medir (Domínguez, 2013; Ferrando *et al.*, 2022). Para medir los reactivos se obtiene un puntaje –grado de presencia de la conducta en un individuo– que se compara con una tabla de normas o baremo –grado de presencia normal para una población– para poder caracterizar el constructo en la persona a quien se le aplica un test (Tovar, 2007).

Ahora bien, los constructos medidos en psicología son, por naturaleza, influenciados por un sistema complejo de variables que suelen ser contingentes al contexto particular donde ocurre un comportamiento (Leong *et al.*, 2016); lo que necesariamente implica que no todo constructo psicológico se presenta en el mismo grado y/o cualidad a nivel general, aunque sus procesos básicos sean compartidos entre contextos. Por ello, para poder medir objetiva y acertadamente un constructo en

un contexto específico, el instrumento empleado debe adaptarse culturalmente al lugar de aplicación (Chahín-Pinzón, 2014; Sánchez-Villena & De La Fuente-Figuerola, 2020). Esto se hace porque las particularidades contextuales modifican la expresión del constructo según la cultura; se reconoce así la importancia de esta, al adaptar y desarrollar instrumentos de evaluación psicométrica (Fernández *et al.*, 2010).

Un proceso completo de adaptación implica varias fases, que en su seguimiento radica la garantía de que el instrumento es ajustado al contexto de interés. Al respecto, Muñiz *et al.*, (2013) proponen los siguientes pasos: permisos con los autores, estudio de las características del constructo en la población objetivo, traducción y revisión y evaluación de la versión en el idioma a trabajar, revisión de las condiciones culturales, una prueba piloto (reacciones a la prueba, errores que los mismos evaluados detecten en formato o contenido), análisis de propiedades métricas (equivalencia de la estructura factorial, fiabilidad, entre otras), estándar de aplicación, interpretación de puntajes (baremación), y documentación de la nueva versión.

Sin embargo, a pesar de seguir estos pasos, es importante dar cuenta que estos no son una fórmula infalible para la adaptación psicométrica; se deben tener en cuenta variables sociales y culturales, que pueden impedir la adaptación de una prueba a cierto contexto (Greiff & Iliescu, 2017). Al adaptar pruebas a distintos contextos, también se pueden encontrar barreras lingüísticas en las pruebas, aun cuando estén en un mismo idioma (Zeinoun *et al.*, 2022). Esto ocurre porque el lenguaje es un sistema de significados construido de forma conjunta por un grupo, lo que resulta en la aparición de expresiones y marcos interpretativos particulares de cada cultura no generalizables a otras (Lira & Caballero, 2020). Lo anterior impacta en las pruebas, ya que son redactadas con un contexto lingüístico específico al lugar y propósito de creación del test. En consecuencia, los ítems no siempre son capaces de representar el constructo en su totalidad; incluso, a veces, resultan incomprensibles o son interpretados de forma incorrecta (Elosua *et al.*, 2014). Esto provoca mediciones imprecisas o erróneas, cuando

no se hace un proceso de adaptación riguroso y contextual; incluso, aun cuando se despliegue un proceso de adaptación cultural esto no garantiza la calidad métrica de la escala.

En conclusión, las escalas de medición y/o evaluación deben ser adaptadas culturalmente, y probadas para dar garantía de su calidad métrica. Sin embargo, este tipo de procesos no son comunes en países latinoamericanos como Colombia (Muñiz *et al.*, 2020; Rueda Serna *et al.*, 2024). En esa línea, un estudio realizado por León (2017) sobre el uso de pruebas por parte de 1226 psicólogos en contexto colombiano encontró que menos del 40 % de ellos refiere conocer evidencias de validez o fiabilidad de dichas escalas, que menos del 20 % conoce estudios de propiedades métricas de la escala que usa para contexto nacional, y que ninguna de las seis pruebas más usadas en Colombia cuentan con baremos a este contexto.

Ahora bien, no siempre la adaptación de un instrumento garantiza una medición exitosa; de ahí la necesidad de poner a prueba el funcionamiento de cada instrumento a utilizar. Precisamente, este estudio tuvo como objetivo presentar una experiencia donde se buscó validar la Escala de Razonamiento Científico propuesta por Drummond y Fischhoff (2017) que, aunque siguió con rigor el proceso de adaptación, el resultado no exhibió la capacidad métrica esperada.

Se contempló esta escala dada su pertinencia social: se parte de la capacidad en el discernimiento de la información verdadera de la información falsa, puesto que, en la actualidad al estar en un momento de generación continua de información en grandes cantidades, se vuelve imprescindible poder distinguir la veracidad de estas (Drummond & Fischhoff, 2017). Un ejemplo de la importancia del razonamiento científico en momentos críticos fue el inicio de la pandemia de Covid19, donde estudios como el de Čavojová y Ballová (2022) encuentran que las personas con un mejor razonamiento científico sentían menos temor a los conflictos de información presentados masivamente por los medios, y lograban evaluar la situación como menos peligrosa,

en comparación con quienes tenían un menor nivel de razonamiento científico.

Esta escala es particularmente valiosa para ejemplificar un caso de diferencias contextuales importantes entre el lugar de desarrollo y el de adaptación. Dadas las características de esta prueba, el nivel de rendimiento académico y competencias intelectuales pueden relacionarse con el razonamiento científico. Por ejemplo, la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2023b) indica que en el mundo el 69 % de los estudiantes alcanzaron el nivel 2 o más de competencia en matemáticas, que en el Reino Unido esta cifra llegó hasta el 76 % y en este territorio el 11 % logró los niveles superiores de la competencia (niveles 5 o 6), mientras que en Colombia solo el 29 % de los estudiantes alcanzó el nivel 2 o más de competencia en matemáticas y menos del 1 % puntúan en niveles 5 o 6 (OCDE, 2023a). Para ambos países el desempeño es similar al observado en 2018, lo que evidencia constancia en las condiciones culturales y académicas dispares entre ambos contextos (OCDE, 2019a; 2019b). Díaz-Pinzón (2021) encuentra hallazgos similares, lo que indica que Colombia se encuentra por debajo del promedio del conjunto de países de la OCDE en el componente de matemáticas.

La brecha en el presupuesto para investigación, así como las falencias en la educación y divulgación científica observada en Latinoamérica contribuyen a dichas diferencias contextuales (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO], 2015; 2016). El informe publicado por la Red Iberoamericana de Indicadores de Ciencia y Tecnología (RICYT, 2024) expone que la inversión en investigación y desarrollo de los países latinoamericanos fue del 0,56 % en 2022, representado en su mayoría por Brasil, México y Argentina. Para Colombia se da un crecimiento en la inversión desde 2015, pero para 2020 representó únicamente el 2,5 % de la inversión iberoamericana en investigación y desarrollo. Teniendo esto en cuenta, el proceso de validación del SRS ilustra un caso de validación que requiere adaptaciones para probar su funcionamiento dadas las diferencias contextuales, lo que lleva a cuestionar si todos los instrumentos psicométricos conservarán sus propiedades

métricas tras un proceso juicioso de adaptación, o si existen diferencias contextuales que resulten en la ineficacia de algunos instrumentos.

2. MÉTODO

2.1. Tipo de estudio

Esta investigación es entendida como un estudio instrumental (Montero & León, 2007), dado que perseguía el proceso de adaptación y revisión de las propiedades métricas; en este caso de la escala de razonamiento científico.

2.2. Participantes

Se contó con la participación de 264 colombianos, con edades entre los 18 y los 74 años (M: 30,81; D.T.: 13,96), siendo 158 mujeres (59,84 %), 105 hombres (39,77 %) y un caso único donde no se reportó sexo, obtenidos por un muestreo no probabilístico tipo bola de nieve (Hernández & Mendoza, 2018). Las características sociodemográficas de los participantes se ven en la Tabla 1. Se toman como criterios de inclusión que los participantes tuviesen ciudadanía y residencia colombiana; como criterios de exclusión que hubiese presencia de algún diagnóstico clínico y la minoría de edad, descartados en la encuesta inicial realizada. El número de participantes se considera suficiente tanto por entenderse

una escala con un modelo SEM simple y con pocos indicadores (Vargas & Mora-Esquivel, 2017).

2.3. Instrumento

La Escala de Razonamiento Científico (SRS), elaborada por Drummond y Fischhoff (2017), se encarga de medir la habilidad para evaluar la evidencia científica desde la metodología científica y la filosofía, comprendiendo el análisis descriptivo del conocimiento y las habilidades de razonamiento científico que pueda tener una persona. La escala cuenta con once (11) ítems que se responden de forma autoaplicada, y que dan información sobre una situación y piden responder verdadero o falso al lector de esta. Es decir, se presenta un contexto por reactivo que permite a quien contesta la prueba entender la pregunta e indicar si la respuesta a esta es verdadera o falsa (Drummond & Fischhoff, 2017). Con relación a la validez de la prueba original de Drummond y Fischhoff (2017), esta tuvo un α de Cronbach de 0,70, un solo factor de distribución de los ítems y, de acuerdo a un AFC, índice de error RMSEA de 0,078, e índice de ajuste comparativo CFI de 0,90. Los ítems traducidos utilizados para el estudio posterior a los ajustes de la prueba piloto se muestran en la Tabla 2.

Tabla 1.

Distribución de participantes

NIVEL ACADÉMICO	F	%	OCUPACIÓN	F	%
Secundaria incompleta/ en curso	3	1,136	Ama de casa	6	2,273
Secundaria completa	3	1,136	Desempleado	5	1,894
Bachillerato incompleto/ en curso	1	0,379	Empleado	87	32,955
Bachillerato completo	40	15,152	Estudiante universitario	113	42,803
Técnico incompleto/en curso	4	1,515	Pensionado	7	2,652
Técnico completo	13	4,924	Trabajador independiente	35	13,258
Tecnológico incompleto/en curso	2	0,758	Otro	11	4,167
Tecnológico completo	7	2,652			
Pregrado incompleto/ en curso	90	34,091			
Pregrado completo	44	16,667			
Especialización completa	39	14,773			
Maestría incompleta/ en curso	3	1,136			
Maestría completa	12	4,545			
Doctorado completo	3	1,136			
PARTICIPACIÓN COMO INVESTIGADOR			ESTRATO	F	%
No	170	64,394	1	9	3,409
Sí	94	35,606	2	39	14,773
			3	70	26,515
			4	93	35,227
			5	32	12,121
			6	21	7,955

Tabla 2.

Escala de Razonamiento científico de Drummond y Fischhoff (2017), traducida y ajustada para la validación en muestras colombianas.

CONCEPTO	ÍTEM	RESPUESTA CORRECTA
Ciego/Doble ciego	1. En una prueba de sabor, un investigador pone café de marca A en una taza con cinta blanca y café de marca B en una taza idéntica con cinta negra. Un asistente de laboratorio les da a los catadores una de las copas, mientras el investigador observa sus expresiones faciales. ¿Verdadero o falso? El ayudante de laboratorio no debe observar cómo se llenan los vasos.	Verdadero
Causalidad	2. Un investigador descubre que las regiones estadounidenses con parques más grandes tienen menos especies en peligro de extinción. ¿Verdadero o falso? Estos datos muestran que aumentar el tamaño de los parques estatales estadounidenses reducirá la cantidad de especies en peligro de extinción.	
Variables de confusión	3. Un investigador pide a los sujetos que armen un rompecabezas en una habitación fría con una radio a todo volumen o en una habitación cálida sin radio. Los sujetos resuelven el rompecabezas más rápidamente en la habitación cálida sin radio. ¿Verdadero o falso? El científico no puede decir si la radio hizo que los sujetos resolvieran el rompecabezas más lentamente.	Verdadero
Validez de constructo	4. Un investigador en educación quiere medir la habilidad matemática general de una muestra de estudiantes de matemáticas de alto rendimiento. Todos los alumnos han tomado clases de geometría y precálculo. ¿Verdadero o falso? El investigador en educación puede medir la habilidad matemática general dando a los estudiantes una prueba de geometría.	Falso
Grupo de control	5. Dos científicos prueban una crema contra el acné, en adolescentes con acné. El científico A quiere dar la crema a todos los adolescentes del estudio. El científico B quiere dar la crema a la mitad de los adolescentes y dar una crema sin ingredientes contra el acné a la otra mitad. ¿Verdadero o Falso? Ambas formas de probar la crema son igualmente adecuadas.	Falso
Validez ecológica	6. Un investigador hace que un grupo de sujetos participe en un juego competitivo. El objetivo de cada sujeto es ganar dinero comprando y vendiendo fichas. A los sujetos se les paga una tarifa fija por participar en el experimento. ¿Verdadero o Falso? El investigador puede afirmar con seguridad que el comportamiento en el experimento refleja el comportamiento de compra y venta de la vida real.	Falso
Historia	7. Se encuesta a una muestra de estadounidenses, seleccionada al azar, sobre la enfermedad A antes y después de una campaña mediática de 6 meses sobre esta enfermedad. A mitad de la campaña mediática, una celebridad famosa muere de la Enfermedad A. Los datos de la encuesta indican que el conocimiento de la Enfermedad A es mayor después de la campaña. ¿Verdadero o Falso? Es posible que la campaña en los medios no haya aumentado el conocimiento de la enfermedad A.	Verdadero
Maduración	8. Los sujetos de un experimento deben presionar un botón cada vez que un punto azul parpadea en la pantalla de su computadora. Al principio, la tarea es sencilla para los sujetos. Pero a medida que continúan realizando la tarea, cometen más y más errores. ¿Verdadero o Falso? El punto azul debe parpadear con mayor rapidez a medida que avanza la tarea.	Falso
Asignación aleatoria a la condición	9. Los investigadores quieren ver si una intervención sanitaria ayuda a los niños en edad escolar a perder peso. Los niños en edad escolar se clasifican en un grupo de intervención o control. ¿Verdadero o Falso? Los investigadores deberían asignar a los niños con sobrepeso al grupo de intervención.	Falso
Confiabilidad	10. Un investigador desarrolla un nuevo método para medir la tensión superficial de los líquidos. Este método es más consistente que el método anterior. ¿Verdadero o Falso? El nuevo método también debe ser más preciso que el método anterior.	Falso
Sesgo de respuesta	11. Dos investigadores están desarrollando una encuesta para medir los sentimientos de los consumidores sobre el servicio al cliente. El investigador A quiere que los clientes califiquen su acuerdo con la afirmación "Estoy satisfecho con el servicio al cliente" en una escala de 5 puntos, donde 1= totalmente de acuerdo y 5= totalmente en desacuerdo. El investigador B quiere que los clientes califiquen el servicio al cliente en una escala de 5 puntos, donde 1= nada insatisfecho y 5= muy insatisfecho. ¿Verdadero o Falso? Estas preguntas son igualmente buenas para medir cómo se sienten los consumidores acerca del servicio al cliente.	Falso

2.4. Procedimiento

Para adaptar el SRS, se siguieron los pasos recomendados por [Muñiz et al., \(2013\)](#): primero se solicitó el consentimiento de los autores origi-

nales ([Drummond & Fischhoff, 2017](#)) para adaptar su prueba al contexto colombiano, que dieron su aprobación para este proceso. Luego, la prueba se tradujo al español mediante un traductor certi-

ficado de inglés. Posteriormente, se sometieron los reactivos traducidos a revisión por jueces expertos, a quienes se les solicitó que calificaran tanto cuantitativa como cualitativamente los ítems. Concluida esta fase, se tomó la decisión de mantener todos los ítems con algunas correcciones para mejorar la claridad y que fueron sugeridas por los jueces. Después, la escala se incluyó en un formulario de Google Forms, estableciendo como criterio obligatorio contestar todos los reactivos para el envío del formulario, y se desarrolló una prueba piloto donde se les pidió a los participantes responder la prueba e indicar si los reactivos eran comprensibles y si se requería algún ajuste. Finalmente se aplicó la prueba a la muestra general posterior a los ajustes (este cuestionario tenía una introducción a la investigación, consentimiento con riesgos, beneficios, objetivo del estudio, y voluntariedad de participar y de retirarse sin limitación o consecuencia alguna). Tras estos pasos, se llevó a cabo la aplicación general, se elaboró la base de datos en Microsoft Excel 2021 y se desarrollaron las pruebas para evaluar las propiedades métricas como se describe en siguiente apartado. En ambos casos de aplicación la prueba fue autoaplicada.

2.5. Enfoque analítico

Se desarrolló un análisis factorial confirmatorio para confirmar la estructura unidimensional previamente encontrada en el estudio inicial (Drummond & Fischhoff, 2017). La bondad de ajuste se evaluó a través de los siguientes indicadores: Raíz del Error Cuadrático Medio de Aproximación (RMSEA), Raíz Cuadrada Media Estandarizada del Residual (SRMR), Índice de Ajuste Comparativo (CFI), Índice de Tucker-Lewis (TLI), Índice de ajuste no normalizado de Bentler-Bonett (NNFI), Índice de ajuste normalizado de Bentler-Bonett (NFI). Para fines

interpretativos se asume un ajuste aceptable del modelo cuando el CFI, TLI, NNFI y NFI presentan valores superiores a 0,90, y el RMSEA y SRMR inferiores a 0,08 (Hu & Bentler, 1999; Hair *et al.*, 2014; Samperio-Pacheco, 2019).

En razón del bajo nivel de ajuste, se aplicó un análisis factorial exploratorio con extracción de mínimos cuadrados ponderados –WLS– y basado en matrices policóricas, por el tipo de reactivos y la presunción de factores interrelacionados (Lloret-Segura *et al.*, 2014), con una depuración de reactivos para encontrar una estructura coherente. También, se aplicaron indicadores de ajuste y separación desde teoría de respuesta al ítem bajo el modelo de Rasch, debido a que la prueba era calificada en modalidad de aciertos. Finalmente, se obtuvieron coeficientes de fiabilidad (alfa de Cronbach, Omega de McDonald y Greatest Lower Bound) con la muestra original y se realizó un bootstrapping a 1000 participantes, para omitir el sesgo por el tamaño muestral. Los análisis realizados desde el modelo de Rasch se ejecutaron desde Winsteps 3.68.2 y los demás por medio del software estadístico JASP versión 17.1.

3. RESULTADOS

Inicialmente, se revisaron los resultados obtenidos en el proceso de validación por jueces respecto a cada reactivo traducido y en cuanto a la prueba en general, en torno a la claridad, coherencia y relevancia –véase tabla 3–. El Índice de validez de contenido general resultó muy alto (IVC=0,94); asimismo todos los reactivos puntuaron por encima del valor mínimo aceptado (0,7) en todas las dimensiones evaluadas, por lo que ningún ítem fue descartado.

Tabla 3.

Índice de validez de contenido

ÍTEM	1	2	3	4	5	6	7	8	9	10	11
Claridad	1,00	0,83	1,00	0,83	1,00	1,00	0,83	1,00	1,00	1,00	1,00
Coherencia	1,00	0,83	1,00	1,00	1,00	1,00	0,83	0,83	0,83	0,83	1,00
Relevancia	1,00	0,83	1,00	1,00	1,00	1,00	1,00	0,83	0,83	0,83	1,00

Posteriormente, se realizaron una serie de análisis de estructura interna y confiabilidad, cuyos resultados se muestran a continuación. Puesto que se trataba de un proceso de adaptación, primero se realizó un análisis factorial confirmatorio (AFC) para 1 factor como proponía la prueba original, previa revisión de los principios de tamaño muestral. En la Tabla 4 se ven los resultados, encontrando un bajo nivel de ajuste para una estructura unidimensional como la planteada en el estudio original.

Tabla 4.

Índices de ajuste del AFC

ÍNDICE	VALOR		
RSMEA	0,042		
SMRM	0,057		
Índice de Ajuste Comparativo (CFI)	0,726		
Índice de Tucker-Lewis (TLI)	0,658		
Índice de ajuste no normalizado de Bentler-Bonett (NNFI)	0,658		
Índice de ajuste normalizado de Bentler-Bonett (NFI)	0,502		
Índice de ajuste normalizado de parsimonia (PNFI)	0,402		
Índice de ajuste relativo de Bollen (RFI)	0,378		
Índice de ajuste incremental de Bollen (IFI)	0,761		
Índice de no centralidad relativa (RNI)	0,726		
CONTRASTE CHI-CUADRADO	VALOR	GL	P
Modelo	54.370	34	0.015

Debido a lo anterior, se llevó a cabo un análisis factorial exploratorio con el fin de revisar posibles agrupaciones viables de los ítems. Los resultados del primer análisis crudo se observan en la tabla 5 donde se muestra la separación en cuatro factores iniciales y descartando los reactivos 5, 9 y 11. Aunque la prueba de esfericidad de Bartlett entregó un valor aceptable ($p = >0,001$), cabe aclarar que esto se llevó a cabo a pesar de haber encontrado valores poco aceptables para el KMO ($0,555 < 0,7$), así como coeficientes de confiabilidad inferiores al mínimo aceptable ($\alpha = 0,460$; $\omega = 0,457$; $GLB = 0,610$).

Tras remover el ítem 11 (tabla 5), no hubo cambios significativos en los coeficientes de verificación de supuestos ($KMO = 0,54$, Bartlett $p = 0 > 0,001$). Sin embargo, los factores se reducen a 3, con valores extremos de unicidad en los factores 1 y 3 y excluyendo los reactivos 2, 3, 7 y 10, causando la

Tabla 5.

Análisis factorial exploratorio

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	UNICIDAD
Item 8	0,569				0,675
Item 2	0,517				0,770
Item 6	0,454				0,677
Item 1		0,963			0,047
Item 4			0,719		0,465
Item 3			0,424		0,791
Item 10			0,416		0,754
Item 7				0,715	0,514
Item 5					0,783
Item 9					0,678
Item 11					0,836

Nota: El método de rotación aplicado es promax

Tabla 6.

Análisis factorial exploratorio tras depuración de reactivos.

	FACTOR 1	FACTOR 2	FACTOR 3	UNICIDAD
Item 4	0,965			
Item 9		0,576		
Item 8		0,547		
Item 5		0,441		
Item 6		0,44		
Item 1			0,908	
Item 2				
Item 3				
Item 7				
Item 10				

Nota: El método de rotación aplicado es promax.

exclusión en conjunto de más del 60 % de los ítems originales con solo 4 restantes –Tabla 6–. Además, los índices de confiabilidad también estuvieron bajo los niveles mínimos aceptables ($\alpha = 0,424$; $\omega = 0,422$; $GLB = 0,582$).

Tras la nueva depuración de reactivos –Tabla 7–, los cuatro reactivos restantes se ajustaron en una estructura bifactorial ($KMO = 0,466$; Bartlett $p = 0 > 0,001$), donde solo cargaron el ítem 8 y el 3, cada uno siendo el único reactivo de su respectiva dimensión, con reducciones significativas en los coeficientes de confiabilidad ($\alpha = 0,165$; $\omega = 0,457$; $GLB = 0,327$).

Tabla 7.

AFE post segunda depuración

	FACTOR 1	FACTOR 2	UNICIDAD
Item 8	1,017		0,004
Item 3		1,024	0,005
Item 2			0,922
Item 4			0,914

Nota: El método de rotación aplicado es promax

A partir de esto, se depuraron los reactivos 2 y 4, lo que deriva en una estructura unifactorial con dos reactivos, el 8 y el 3. Aunque el KMO aumentó respecto a la depuración anterior (KMO= 0,5), la prueba de esfericidad de Bartlett indicó bajo ajuste de los datos para realizar análisis factoriales ($p= 0,229$), lo que es esperable debido a la poca cantidad de reactivos. Asimismo, como se observa en la tabla 6, los índices de confiabilidad fueron muy bajos ($\alpha= 0,089$; $\omega= 0,144$; $GLB= 0,089$) y la correlación entre los únicos 2 ítems que se ajustaron a la estructura factorial no resultó significativa (Rho de Spearman = 0,046; $p=0,453$).

Puesto que la prueba en cuestión contiene respuestas dicotómicas con posibles respuestas correctas –similar a una prueba de conocimiento–, también se revisó su estructura interna con modelos de Teoría de Respuesta al Ítem a partir de modelos de Rasch. Sin embargo, en este caso tampoco se encontraron indicadores aceptables respecto al funcionamiento de la prueba. Si bien el índice de separación fue aceptable ($2,73 \approx 3$), los índices arrojados por el análisis presentaron un bajo nivel de ajuste, como se expone en la Tabla 8.

Tabla 8.

Resultados generales y por ítems del análisis con modelos de Rasch

RESULTADOS GENERALES	INFIT	OUTFIT	SEPARACIÓN
Prueba	2,29	3,46	2,73
ITEMS	INFIT	OUTFIT	PT-MEASURE
10	2,20	6,12	-0,06
11	2,21	3,32	0,11
8	2,38	3,56	0,03
1	2,45	3,61	0,01
7	2,56	3,69	-0,03
2	2,29	3,32	0,08
4	2,16	3,03	0,14
6	2,23	3,25	0,10
3	2,54	4,18	-0,06
5	2,33	4,07	0,02
9	2,24	4,03	0,05

Los índices de ajuste generales y por ítem tampoco presentaron valores aceptables. Se encontró un índice de ajuste respecto al modelo ideal –infit superior a 1,7–, lo que indica que la prueba no se ajusta al modelo unidimensional, presupuesto necesario para los reactivos dicotómicos, y que la variable medida en cuestión no es clara (Linacre, 1999). En cuanto al índice de ajuste externo, que revisa el número y estructura de respuestas anómalas, también excedió el valor máximo aceptable $-3,46 > 1,7$ –, lo que indica que las respuestas del test son demasiado variables y no se ajustan al modelo predictivo inferido de la escala, lo que implica una baja capacidad de predicción.

Los valores de los índices de ajuste interno y externo por reactivo también exceden los valores aceptables, sugiriendo en conjunto que la prueba no mide con precisión la variable deseada, y teniendo en cuenta la dispersión de las respuestas es posible concluir que la variable tampoco se mide en su continuo. Además, las medidas de correlación biserial por reactivo fueron muy bajas, todas inferiores a 0,15 siendo el mínimo aceptable 0,3, lo que sugiere que los reactivos no miden el mismo constructo. En conjunto, los índices de ajuste extraídos del análisis de modelos de Rasch indicaron que la prueba, pese a tener un índice de separación apenas suficiente, no mide la variable objetivo apropiadamente.

4. DISCUSIÓN

El objetivo de esta investigación fue, inicialmente, adaptar la escala de razonamiento científico (SRS) al marco colombiano. Sin embargo, tras realizar los análisis de validez y confiabilidad, se encontró que la prueba no es aplicable al contexto colombiano, esto teniendo en cuenta los resultados de los análisis factoriales confirmatorios y exploratorios, donde en ambos casos la prueba mostró bajos niveles de ajuste. Asimismo, al revisar la estructura interna de la prueba mediante modelos de Rasch, se obtuvieron resultados similares, encontrando una baja capacidad para medir apropiadamente la variable objetivo, contrario a lo encontrado en el estudio original. En cuanto a los coeficientes de confiabilidad, tampoco se encontraron valores suficientes ($\alpha=0,460$; $\omega= 0,457$; $GLB=0,610$) para afirmar que la prueba es confiable; esto a pesar de buscar eliminar posibles fuentes de error como incomprensibilidad mediante la traducción oficial,

la revisión por jueces y la prueba piloto. A partir de lo anterior, se modifica el objetivo del estudio para exponer un caso de adaptación cuyos resultados indicaron un instrumento con limitaciones métricas para Colombia a pesar de la rigurosidad del proceso dado y el seguimiento de lineamientos internacionales. El aporte de esta investigación es justamente indicar que algunos instrumentos en procesos de adaptación no funcionan para todos los contextos.

Concretamente, adaptar una escala de razonamiento científico puede enfrentar dificultades en términos de adaptaciones culturales debido a la falta de acceso al conocimiento científico y las diferencias en el contexto sociocultural. Al respecto, [Elosua et al., \(2014\)](#) ya referían que los constructos psicológicos pueden ser afectados por condiciones culturales o sociales; precisamente allí radica la necesidad de adaptar y probar los instrumentos en distintos contextos. Al respecto, la [UNESCO \(2015\)](#) halla que solo el 1 % de la población en el mundo se dedica a la investigación científica, repartándose solo el 8 % del total en América Latina, mientras que el 72 % está en China, la Unión Europea, Japón, Rusia y Estados Unidos, lo que pondría de manifiesto políticas gubernamentales diferentes en torno a la ciencia y la investigación. En concreto, las revisiones de la [UNESCO \(2016\)](#) indican falencias en los programas de Latinoamérica respecto a la educación en ciencia y su divulgación, así como bajos niveles de interés por parte de la población general en estos temas. Además, se evidencia una inversión muy reducida en investigación y desarrollo en Latinoamérica ([RICYT, 2024](#)). En conjunto, esto podría recaer en las diferencias culturales que llevan al no mantenimiento de la calidad métrica del SRS.

Igualmente, el modo de interpretación de las preguntas y los procesos de pensamiento utilizados para responder también pueden tener influencia en los resultados. Puesto que ambos están mediados por variables culturales y de contexto, realidades como la descrita en el párrafo anterior pueden llevar

a diferentes formas de comprender el contenido de las preguntas y, por ende, modificar los procesos de pensamiento utilizados para responderlas ([Elosua et al., 2014](#)). A nivel de desempeño, los resultados constantes en la prueba del Programa para la Evaluación Internacional de Alumnos (PISA) tanto para 2018 y 2022 ([OCDE, 2019a; 2019b; 2023a; 2023b](#)) y los hallazgos de [Díaz-Pinzón \(2021\)](#) indican que hay una brecha importante entre los países donde se encuentra probada la prueba y Colombia en cuanto a desempeño académico general, donde se evidencian niveles de rendimiento general muy inferiores a los de Inglaterra. Por ende, los niveles de razonamiento científico podrían no ser equivalentes entre ambos países. Así, estas diferencias culturales y contextuales pueden afectar la validez y confiabilidad de la adaptación transcultural del SRS.

5. LIMITACIONES Y RECOMENDACIONES

En cuanto a las limitaciones de este estudio, hay dos aspectos a considerar: en primer lugar, aunque se logró alcanzar el tamaño mínimo de muestra recomendado por algunos autores ([Lloret-Segura, 2014](#)) —que es de 200 participantes cuando hay entre 3 y 4 ítems por dimensión y comunalidades entre 0,4 y 0,7—, no se puede garantizar que sea suficiente para evaluar la adaptabilidad del instrumento; aun así, tras realizar el *bootstrap* a 1000, se encuentran resultados poco alentadores respecto a los coeficientes de confiabilidad. En segundo lugar, existe otra limitación importante frente al uso del modelo Rasch para el análisis, ya que su validez y confiabilidad pueden verse comprometidas cuando se emplea con muestras inferiores a 500 participantes. Aun así, este sesgo solo es grave cuando $n > 200$, lo que no ocurrió en este estudio ([Müller, 2020](#)). Estas limitaciones deben tenerse en cuenta al interpretar los resultados del estudio y pueden afectar la generalización y la fiabilidad de las conclusiones obtenidas. Otras limitaciones responden al muestreo no probabilístico, que atenta contra la generalización de estos resultados.

Los resultados de este estudio sugieren algunas investigaciones futuras, entre ellas podría desarrollarse una nueva adaptación del SRS a otro contexto ajeno a Colombia, esto con el propósito de evaluar el comportamiento de la prueba en culturas distintas, por ejemplo, otros países de Latinoamérica. De esta forma sería posible confirmar si las dificultades métricas evidenciadas son generalizables o específicas de la muestra utilizada, y si son causadas por variables culturales. Además, se plantea realizar futuras adaptaciones del constructo con diferentes escalas—que sean culturalmente equitativas— o la creación de una escala de razonamiento científico específica para el contexto colombiano. En torno a tipos de análisis, se recomienda ampliar a propuestas de ESEM (exploring structural equation modeling) que pueden referir con mayor eficiencia las estructuras internas de las pruebas, y los análisis de invarianza factorial como herramienta de verificación de alteraciones estructurales entre grupos.

En síntesis, este estudio ejemplifica el problema de aplicar pruebas traducidas sin asegurar su validez y confiabilidad de antemano, puesto que sus propiedades métricas pueden no funcionar de la forma esperada, y como consecuencia llevar a resultados sesgados en los procesos de investigación o evaluación psicológica en los que se empleen y muestra cómo a pesar de un proceso sistemático y riguroso no todo instrumento llega a funcionar en todos los contextos. Por ende, para garantizar la precisión de los procesos de medición de constructos psicológicos en todo contexto, es necesario desarrollar nuevos instrumentos funcionales en su intermediación cultural concreta, así como realizar las debidas adaptaciones y pruebas de funcionamiento métrico de los tests que vayan a utilizarse en dicho contexto.

6. HIGHLIGHTS

- Los instrumentos de medición y evaluación psicológica deben adaptarse al contexto cultural donde van a ser utilizados, y debe

probarse la calidad métrica (evidencias de validez y fiabilidad).

- Es un error asumir que, porque esté traducido al idioma donde va a ser utilizado, una escala siempre funciona en este nuevo contexto.
- La Escala de razonamiento científico (SRS) no tiene un adecuado funcionamiento métrico para Colombia, posiblemente porque hay importantes diferencias culturales entre este país y el de origen de la escala.

Conflicto de interés: Declaramos que no existieron conflictos de intereses durante la realización de esta investigación. Ninguna circunstancia personal, académica o financiera influyó en el desarrollo o los resultados del estudio.

Agradecimientos: Agradecemos a la Universidad de La Sabana por su apoyo y financiación a este proyecto, a los autores originales de la escala por su colaboración en el proceso de validación, a los jueces expertos (Daniela Bohórquez, Diego Efrén Rodríguez, Víctor Oswaldo Gamboa, Luis Enrique Prieto y Julien Dampuré) por sus valiosos aportes en la revisión del instrumento, y a cada uno los participantes que contribuyeron con el proceso de validación por su aporte a la construcción de nuevo conocimiento.

Descargo de responsabilidad: Las opiniones y conclusiones presentadas en este trabajo son propias de los autores y no reflejan necesariamente la posición institucional ni los criterios de la universidad

Financiamiento: Este estudio es derivado del proyecto “Factores de riesgo y protección asociados a conductas de riesgo y problemas que afectan la salud mental en niños y adolescentes”, con código PSIPHD-4-2023, y avalado por la Subcomisión de Investigación y Ética de la Facultad de Ciencias del comportamiento de la Universidad de La Sabana, y financiado por la Universidad de La Sabana.

REFERENCIAS

- Čavojová, V., Šrol, J., & Ballová Mikušková, E. (2022). How scientific reasoning correlates with health-related beliefs and behaviors during the COVID-19 pandemic? *Journal of health psychology*, 27(3), 534-547. <https://doi.org/10.1177/1359105320962266>
- Chahín-Pinzón, N. (2014). Aspectos a tener en cuenta cuando se realiza una adaptación de test entre diferentes culturas. *Psychologia. Avances de la disciplina*, 8(2), 109-112. <https://www.redalyc.org/pdf/2972/297232756008.pdf>

- Díaz-Pinzón, J. (2021). Análisis de los resultados de la prueba Pisa 2018 en matemáticas para América. *Revista de Investigaciones Universidad del Quindío*, 33(1), 104-114. <https://doi.org/10.33975/riuq.vol33n1.463>
- Domínguez, S. (2013). ¿Ítems Politémicos o Dicotómicos? Un estudio empírico con una escala unidimensional. *Revista Argentina de Ciencias del Comportamiento*, 5(3), 30-37. <https://doi.org/10.32348/1852.4206.v5.n3.5562>
- Drummond, C., & Fischhoff, B. (2017). Development and validation of the scientific reasoning scale. *Journal of Behavioral Decision Making*, 30, 26-38. <https://doi.org/10.1002/bdm.1906>
- Elosua, P., Mujika, J., Almeida, L. & Hermosilla, D. (2014). Procedimientos analítico-razonales en la adaptación de tests. Adaptación al español de la batería de pruebas de razonamiento. *Revista Latinoamericana de Psicología*, 46(2), 117-126. [https://doi.org/10.1016/S0120-0534\(14\)70015-9](https://doi.org/10.1016/S0120-0534(14)70015-9)
- Fernández, A., Pérez, E., Alderete, A., Richaud, M., & Liporace, M. (2010). ¿Construir o adaptar tests psicológicos? Diferentes respuestas a una cuestión controvertida. *Revista Evaluar*, 10(1), 60-74. <https://doi.org/10.35670/1667-4545.v10.n1.459>
- Ferrando, P., Lorenzo-Seva, U., Hernández-Dorado, A. & Muñiz, J. (2022). Decalogue for the factor analysis of test items. *Psicothema*, 34(1), 7-17. <https://www.psicothema.com/pdf/4715.pdf>
- Greiff, S., & Iliescu, D. (2017). A Test Is Much More Than Just the Test Itself. *European Journal of Psychological Assessment*, 33(3), 145-148. <https://doi.org/10.1027/1015-5759/a000428>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis* (7th ed.). Pearson Education Limited.
- Hernández, R. & Mendoza, C. (2018). *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta*. Mc Graw Hill: México.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- León, F. (2017). Evaluación piloto de la calidad técnica de seis de las pruebas psicológicas más usadas en Colombia [Tesis de grado de Maestría]. Universidad Nacional de Colombia. <https://repositorio.unal.edu.co/bitstream/handle/unal/62036/52213783.2017.TESIS.pdf?sequence=4&isAllowed=y>
- Leong, F. T., Bartram, D., Cheung, F., Geisinger, K. F., & Iliescu, D. (2016). *The ITC international handbook of testing and assessment*. Oxford University Press. <https://global.oup.com/academic/product/the-itc-international-handbook-of-testing-and-assessment-9780199356942?cc=us&lang=en&>
- Linacre, J. M. (1999). Understanding Rasch measurement: estimation methods for Rasch measures. *Journal of outcome measurement*, 3, 382-405. <https://pubmed.ncbi.nlm.nih.gov/10572388/>
- Lira, M. & Caballero, E. (2020). Adaptación transcultural de instrumentos de valuación en salud: Historia y reflexiones del por qué, cómo y cuándo. *Revista Médica Clínica Las Condes*, 31(1), 85-94. <https://doi.org/10.1016/j.rmcl.2019.08.003>
- Livia, J., & Ortiz, M. (2014). *Construcción de pruebas psicométricas: Aplicaciones a las ciencias sociales y de la salud*. Universidad Nacional Federico Villarreal. <http://dx.doi.org/10.19083/ridu.10.479>
- 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*, 30(3), 1151-1169. <http://dx.doi.org/10.6018/analesps.30.3.199361>
- Martínez, R. & Villota, H. (2022). La Psicometría. *Revista Sigma*, 18(1), 23-29. <https://revistas.udenar.edu.co/index.php/rsigma/article/view/7951>
- Montero, I. & León O. (2007). A guide for naming research studies in Psychology. *International journal of clinical and health psychology*, 7(3), 847-862. <https://www.redalyc.org/articulo.oa?id=33770318>
- Morucci, M., Foster, M. J., Webster, K., Lee, S. J., & Siegel, D. A. (2024). Measurement That Matches Theory: Theory-Driven Identification in Item Response Theory Models. *American Political Science Review*, 1-19. <https://doi.org/10.1017/S000305542400039X>
- Muñiz, J., Elosua, P. & Hambleton, R. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, 25(2), 151-157. <https://doi.org/10.7334/psicothema2013.24>
- Muñiz, J., Hernández, A., Elosua, P., & Wechsler, S. M. (2020). El uso de los test en España y Latinoamérica: perspectivas actuales y retos futuros. *Anuario Internacional de Revisiones en Psicología*, 1, 67-94. <https://doi.org/10.14635/REVPSY.0.4>
- Müller, M. (2020). Item fit statistics for Rasch analysis: can we trust them?. *Journal of Statistical Distributions and Applications*, 7(1), 1-12. <https://doi.org/10.1186/s40488-020-00108-7>
- OCDE. 2019a. Colombia - Country Note - PISA 2018 Results. *Organización para la Cooperación y el Desarrollo Económicos*. https://www.oecd.org/content/dam/oecd/en/about/programmes/edu/pisa/publications/national-reports/pisa-2018/featured-country-specific-overviews/PISA2018_CN_COL_ESP.pdf
- OCDE. 2019b. PISA 2018 Results. What students know and can do. Volume I. *Organización para la Cooperación y el Desarrollo Económicos*. https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/12/pisa-2018-results-volume-i_947e3529/5f07c754-en.pdf
- OCDE. 2023a. PISA 2022 Results (Volume I and II) - Country Notes: Colombia. *Organización para la Cooperación y el Desarrollo Económicos*. <https://www.oecd.org/en/publications/pisa-2022-re->

sults-volume-i-and-ii-country-notes_ed6fbcc5-en/colombia_dd5f34d9-en.html

OCDE. (2023b). PISA 2022 Results (Volume I and II) - Country Notes: United Kingdom. *Organización para la Cooperación y el Desarrollo Económicos*. https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/united-kingdom_9c15db47-en.html

RICYT. (2024). *El estado de la Ciencia*. Red Iberoamericana de Indicadores de Ciencia y Tecnología. <https://www.ricyt.org/wp-content/uploads/2024/12/El-Estado-de-la-Ciencia-2024.pdf>

Rueda Serna, V., Tejada Castaño, S. V., & Lauritzen Guzmán, E. (2024). Validación de pruebas psicométricas en América Latina: revisión sistemática [Revisión sistemática]. *Universidad Cooperativa de Colombia Sede Villavicencio*. <https://repository.ucc.edu.co/server/api/core/bitstreams/d51f50c0-76ba-4b69-a763-6e15492dca77/content>

Samperio Pacheco, V. M. (2019). Ecuaciones estructurales en los modelos educativos: Características y fases en su construcción. *Apertura*, 11(1), 90-103. <http://dx.doi.org/10.32870/Ap.v11n1.1402>

Sánchez-Villena, A. R. & De La Fuente-Figuerola, V. (2020). Estandarización, adaptación y validación de pruebas psicométricas: diferencias necesarias. *Anales de Pediatría: Publicación Oficial de la Asociación Española de Pediatría (AEP)*, 93(5), 354-354. <http://dx.doi.org/10.1016/j.anpedi.2020.05.014>

Sarabia, C. & Alconero, A. (2019). Claves para el diseño y validación de cuestionarios en Ciencias de la Salud. *Enfermería en Cardiología*, 26(77), 69-73. https://enfermeriaencardiologia.com/media/acfupload/627a2069dc5b2_Enferm-Cardiol.-2019-26-77-Art_1-1.pdf

Tovar, J. (2007). Psicometría: Test Psicométricos, Confiabilidad y Validez. En: A. Quintana & W. Montgomery (Eds.), *Psicología: Tópicos de Actualidad*. (pp. 85-108). UNMSM.

Unesco (2015). *Informe de la Unesco sobre la ciencia, hacia 2030. Resumen ejecutivo*. https://unesdoc.unesco.org/ark:/48223/pf0000235407_spa.page=4

Unesco. (2016). *Educación científica. Resumen ejecutivo*. <https://unesdoc.unesco.org/ark:/48223/pf0000246427>

Vargas, T. & Mora-Esquivel, R. (2017). Tamaño de la muestra en modelos de ecuaciones estructurales con constructos latentes: Un método práctico. *Actualidades Investigativas en Educación*, 17(1), 1-34. <http://dx.doi.org/10.15517/aie.v17i1.27294>

Zeinoun, P., Iliescu, D., & El Hakim, R. (2022). Psychological tests in Arabic: A review of methodological

practices and recommendations for future use. *Neuropsychology Review*, 32(1), 1-19. <https://doi.org/10.1007/s11065-021-09476-6>

Esta obra está bajo: Creative commons attribution 4.0 international license. El beneficiario de la licencia tiene el derecho de copiar, distribuir, exhibir y representar la obra y hacer obras derivadas siempre y cuando reconozca y cite la obra de la forma especificada por el autor o el licenciente.

