

Adaptación cultural al español del Perú y validación psicométrica de un instrumento para medir el alfabetismo científico en estudiantes de medicina

Cultural adaptation to Peruvian Spanish and psychometric validation of an instrument to measure scientific literacy in medical students

Franco Romaní-Romaní^{1,2,a}

¹ Facultad de Medicina, Universidad Nacional Mayor de San Marcos. Lima, Perú.

² Facultad de Medicina Humana, Universidad de Piura. Lima, Perú.

^a Magister en epidemiología. ORCID: <https://orcid.org/0000-0002-6471-5684>

An Fac med. 2024;85(1):34-42./ DOI: <https://doi.org/10.15381/anales.v85i1.26534>.

Correspondencia:

Franco Romaní Romaní
franco.romani@udep.edu.pe

Recibido: 24 de octubre 2023

Aprobado: 26 de febrero 2024

Publicación en línea: 25 de marzo 2024

Conflicto de interés: El autor declara que tiene vínculo laboral con la Universidad de Piura.

Fuente de financiamiento: El estudio fue financiado por la Facultad de Medicina Humana de la Universidad de Piura.

Contribuciones de autoría: Franco Romaní-Romaní conceptualizó el estudio, formuló el diseño metodológico, realizó el análisis formal de los datos, condujo la investigación, realizó la curación de los datos, redactó el primer borrador del artículo, revisó y aprobó la versión final que fue sometida, y realizó la administración del proyecto.

El presente estudio forma parte de la tesis: Romaní F. Efecto de una intervención educativa curricular sobre las habilidades en alfabetismo científico en estudiantes de medicina humana de una universidad de Lima, Perú [tesis de doctorado]. Lima: Facultad de Medicina, Universidad Nacional Mayor de San Marcos; 2023.

Citar como: Romaní-Romaní F. Adaptación cultural al español del Perú y validación psicométrica de un instrumento para medir el alfabetismo científico en estudiantes de medicina. An Fac med. 2024;85(1):34-42. DOI: <https://doi.org/10.15381/anales.v85i1.26534>.

Resumen

Introducción. El alfabetismo científico comprende habilidades para explicar los fenómenos científicamente, evaluar la ciencia e interpretar sus resultados. Para asegurar su correcta medición es necesario disponer de herramientas validadas en diversos idiomas. **Objetivo.** Adaptar y validar, al español del Perú, un instrumento para medir las habilidades en alfabetismo científico en estudiantes de medicina. **Métodos.** La adaptación cultural del *Test of Scientific Literacy Skills* (TOSLS) tuvo seis pasos. Primero, una traducción independiente y simultánea del inglés al español. Luego, síntesis de ambas traducciones y traducción inversa, seguida de la comparación con la versión original y su retrotraducción. Luego, se realizó una prueba piloto y un estudio transversal con estudiantes de medicina en Lima, Perú. El último paso comprendió la evaluación psicométrica utilizando el modelo de Rasch. El instrumento tuvo 28 preguntas de opción múltiple con una única respuesta correcta. **Resultados.** Enrolamos 300 estudiantes. El coeficiente Kuder-Richadson fue 0,742. Los ítems mostraron unidimensionalidad e independencia local (MADaQ3 = 0,054). El ítem 12 fue el más difícil, el 5 resultó el más fácil. Los estadísticos *infit* y *outfit* de los ítems estuvieron en el rango de 0,7 a 1,3 logits. La curva de función de información brindó información con mayor precisión en el nivel de habilidad ubicado entre -1 y 0 logits. Hubo correspondencia entre la dificultad de los ítems y la habilidad en alfabetismo científico. **Conclusión.** La versión en español mantuvo los 28 ítems del TOSLS original y tuvo aceptable consistencia interna. Los ítems tuvieron propiedades aceptables, independientemente de la habilidad en alfabetismo científico de los estudiantes.

Palabras clave: Estudiantes de Medicina; Psicometría; Ciencia, Tecnología y Sociedad; Alfabetización (fuente: DeCS BIREME).

Abstract

Introduction. Scientific literacy comprises skills to explain phenomena scientifically, evaluate science and interpret its results. Validated tools in different languages are needed to ensure their correct measurement. **Objective.** To adapt and validate, to Peruvian Spanish, an instrument to measure scientific literacy skills in medical students. **Methods.** The cultural adaptation of the *Test of Scientific Literacy Skills* (TOSLS) into Spanish had six steps. First, an independent, simultaneous translation from English into Spanish was performed. Then, both translations were synthesised and back-translated, followed by comparison with the original version and back-translation. This was followed by a pilot test and a cross-sectional study with medical students in Lima, Peru. The last step involved psychometric evaluation using the Rasch model. The instrument consisted of 28 multiple-choice questions with only one correct answer. **Results.** 300 students were enrolled. The Kuder-Richadson coefficient was 0.742. The items showed unidimensionality and local independence (MADaQ3 = 0.054). Item 12 was the most difficult, while item 5 was the easiest. The *infit* and *outfit* statistics of the items were in the range of 0.7 to 1.3 logits. The test information function provided information with greater precision at skill levels between -1 and 0 logits. There was a correspondence between item difficulty and scientific literacy ability. **Conclusion.** The Spanish version retained the 28 items of the original TOSLS and had acceptable internal consistency. The items had acceptable properties, independently of students' scientific literacy ability.

Keywords: Students, Medical; Psychometrics; Science, Technology and Society; Literacy (source: MeSH NLM).

INTRODUCCIÓN

El alfabetismo científico (AC) es la capacidad de comprender y participar en discusiones críticas sobre la ciencia y tecnología. Estas habilidades comprenden tres competencias fundamentales: a. explicar científicamente los fenómenos; b. evaluar y diseñar una investigación científica; y, c. interpretar los datos y evidencia científica para llegar a conclusiones científicas adecuadas⁽¹⁾. El AC ha sido medido en muestras representativas de la población⁽²⁾. Sin embargo, también se ha fomentado y medido en estudiantes universitarios de diversas disciplinas⁽³⁻⁵⁾, incluyendo aquellos de ciencias biológicas y la salud⁽⁶⁻⁸⁾.

Uno de los aspectos más destacados en la investigación del AC se centra en el desarrollo de herramientas para su medición. Una revisión sistemática identificó 13 instrumentos para medir el AC⁽⁹⁾. Los instrumentos más utilizados fueron el TBSL (del inglés *Test of Basic Scientific Literacy*)⁽¹⁰⁾, el TOSLS (del inglés *Test of Scientific Literacy Skills*)⁽¹¹⁾, el SLA (del inglés *Scientific Literacy Assessment*)⁽¹²⁾ y el GSLQ (del inglés *Global Scientific Literacy Questionnaire*)⁽¹³⁾. Es importante destacar que, con la excepción del TOSLS, diseñado específicamente para estudiantes universitarios de pregrado en biología, los demás instrumentos están dirigidos al público escolar.

La herramienta TOSLS ha sido usada en 10 de 43 estudios que midieron el AC, estas investigaciones fueron realizadas en Estados Unidos, Brasil, Indonesia, Canadá, Eslovaquia y Nigeria, y por lo tanto, la herramienta ha sido traducido al portugués, indonesio, eslovaco y Nijá (Nigeria)⁽⁹⁾. La herramienta TOSLS está diseñada para medir dos aspectos fundamentales del AC: el primero se refiere a las habilidades de reconocimiento y análisis del uso de los métodos de investigación para producir conocimiento científico; y la segunda, a las habilidades de organización, análisis e interpretación cuantitativa de datos e información científica⁽¹¹⁾.

La mayoría de los estudios previos en grupos de diversas disciplinas científicas usaron el TOSLS original^(4,14-16) o sus versiones traducidas^(17,18). La única evaluación en estudiantes de medicina fue

realizada en la India aplicando la versión original en inglés⁽¹⁹⁾. A nuestro conocimiento, el TOSLS no ha sido objeto de estudios de traducción, adaptación y validación transcultural al español.

La evaluación del AC en estudiantes que emprenden la carrera médica debe ser rigurosa y capaz de medir habilidades como la recuperación e integración de información, el procesamiento y análisis cuantitativo de datos e información numérica, y la capacidad de adquirir y utilizar información de fuentes primarias para respaldar argumentos y conclusiones. Por lo descrito, el objetivo de este estudio fue adaptar y validar, al español del Perú, un instrumento para medir las habilidades en AC en estudiantes de medicina.

MÉTODOS

Ámbito y diseño de estudio

El estudio fue realizado en una escuela de medicina de una universidad privada ubicada en la ciudad de Lima, Perú. El proceso de adaptación transcultural del instrumento TOSLS fue realizado según Sousa V. *et al*⁽²⁰⁾, mientras que la evaluación psicométrica del instrumento fue realizada mediante un diseño transversal (Figura 1).

Descripción del TOSLS

El TOSLS evalúa nueve habilidades de AC agrupadas en dos categorías. La primera agrupa ítems que evalúan como los estudiantes «entienden los métodos de investigación que producen el conocimiento científico», la segunda mide las «habilidades para organizar, analizar e interpretar datos cuantitativos e información científica», cada categoría incluye cuatro y cinco habilidades, respectivamente (Material suplementario 1: tabla 1). Las habilidades son medidas con 28 ítems correspondientes a preguntas de opción múltiple con mejor respuesta única. Las preguntas tuvieron cuatro opciones, una correcta y tres distractores.

El TOSLS fue validado en una muestra de 296 estudiantes universitarios inscritos en el curso introductorio de biología de una universidad en los Estados Unidos de Norteamérica. Las propiedades psicométricas fueron evaluadas al inicio y al finalizar un semestre académico. La confiabilidad interna estimada con la fórmu-

la de Kuder-Richarson encontró un valor de confiabilidad del 0,731 y 0,748 en la aplicación pre y post semestre, respectivamente. El análisis factorial exploratorio con el método de rotación Varimax sugirió que un factor explicó mejor la varianza de los datos⁽¹¹⁾.

Traducción y adaptación cultural

El proceso de adaptación cultural de seis pasos es descrito en la figura 1. En el paso cinco se realizó el piloto en 24 estudiantes de medicina del séptimo año (tasa de respuesta: 88,9%). El piloto fue realizado mediante la aplicación del cuestionario impreso. El objetivo fue evaluar la claridad de las instrucciones, la redacción de la viñeta, de la pregunta introductoria y las opciones de cada ítem del instrumento. A cada participante se le pidió que evaluará la claridad del ítem utilizando una escala del 1 al 10 (1: menor claridad a 10: mayor claridad). También se solicitó a los estudiantes que ofrecieran sugerencias para reformular los ítems que consideraran poco claros.

Entre los participantes en el piloto, la mediana del puntaje fue 18,5 con un percentil 25 y 75 de 13,2 y 22,7, respectivamente. Los ítems 5 y 22 fueron respondidos correctamente por todos los estudiantes, el ítem 24 fue respondido correctamente por el 47,4% de estudiantes. Respecto a la calificación de claridad de cada ítem, se usó una escala del 1 (poco claro) a 10 (muy claro); el ítem 1 tuvo una mediana de 8,5, el ítem 11 de 9,5; los demás ítems obtuvieron una mediana de 10. El tiempo de entrega para completar el instrumento varió de 42 (primero en entregar) a 49 minutos (último). En el último paso, la versión prefinal en español del TOSLS fue sometido a una validación psicométrica.

Población de estudio y muestra

La validación psicométrica se realizó con estudiantes de medicina del primer al sexto año. Los criterios de selección incluyeron estar matriculado en el segundo semestre de 2023 y proporcionar consentimiento informado. Fueron excluidos aquellos que respondieron 13 o menos ítems del instrumento y aquellos que entregaron el instrumento en menos de 15 minutos. La administración de la prueba fue del 18 de agosto al 7 de setiembre del 2023.

Tabla 1. Índice de dificultad y estadísticos de ajuste de los ítems de la prueba para medir el alfabetismo científico (TOSLS en español), ordenados por nivel de dificultad.

Número de ítem	Dificultad del ítem ^a	Error estándar	Índice Infit MNSQ ^b	Índice Outfit MNSQ ^b
12	1,52	0,15	1,10	1,22
14	0,93	0,13	1,09	1,15
15	0,23	0,12	1,12	1,16
24	0,05	0,12	1,02	1,01
18	-0,06	0,12	0,97	0,97
10	-0,16	0,12	1,04	1,05
11	-0,25	0,12	1,11	1,14
2	-0,36	0,12	1,00	1,00
20	-0,39	0,12	0,94	0,93
19	-0,42	0,13	0,95	0,93
28	-0,45	0,13	1,02	1,03
13	-0,48	0,13	1,04	1,04
16	-0,68	0,13	0,95	0,92
4	-0,74	0,13	0,96	0,94
17	-0,78	0,13	1,05	1,06
21	-0,86	0,13	1,00	1,01
3	-0,95	0,13	0,98	0,95
7	-1,05	0,14	1,00	0,99
8	-1,05	0,14	0,93	0,87
6	-1,15	0,14	0,88	0,81
25	-1,28	0,14	1,04	1,07
26	-1,33	0,14	1,08	1,10
1	-1,52	0,15	0,95	0,93
23	-1,63	0,15	0,96	0,91
27	-1,71	0,16	0,96	0,91
9	-2,10	0,18	0,94	0,83
22	-2,58	0,21	0,99	0,85
5	-2,77	0,23	0,93	0,67

^aPuntajes más negativos indican ítem más fáciles

^bRangos aceptables de estadísticos de ajuste outfit y infit: 0,70-1,30 logits.

Índices MNSQ > 1,30 (ítems inconsistentes o que no miden el constructo subyacente)

Índices MNSQ < 0,70 (ítems redundantes)

El mínimo tamaño de muestra fue determinado en 107 sujetos utilizando un alfa de Cronbach de 0,748 —como aproximación al coeficiente de Kuder-Richarson—, con una precisión relativa del 10%, un nivel de significancia de 5%, 28 ítems y pérdida esperada del 10%. Sin embargo, para este estudio se consideró 10 sujetos por ítem para realizar estimaciones válidas de los estadísticos psicométricos (280 estudiantes) ⁽²¹⁾.

La muestra estuvo constituida por todos los estudiantes que cumplieron los

criterios de selección. Se invitó a toda la población de estudio para asegurar el tamaño muestral mínimo necesario. La población de estudio estuvo constituida por 497 sujetos: 192 en primer año, 92 en segundo, 51 en tercero, 56 en cuarto, 60 en quinto y 26 en sexto año.

Recogida de datos

La aplicación del instrumento fue presencial con un cuestionario impreso. Previamente, se proporcionó una explicación sobre el estudio y se recopiló el

consentimiento informado por escrito de los estudiantes. Se brindó 50 minutos, cada estudiante recibió una ficha óptica para completar su respuestas. El uso de las fichas ópticas permitió automatizar la captura de los datos. La aplicación del cuestionario fue realizada por el autor durante el horario de clases.

Análisis de datos

Se realizó el análisis descriptivo (mediana, percentil 25 y 75) del número de respuestas correctas para toda la muestra. Para estimar el índice de dificultad y de discriminación de cada ítem, se aplicó la teoría clásica de la prueba (TCP). El índice de dificultad se calculó como el porcentaje de estudiantes que respondieron correctamente a cada ítem, considerando como nivel aceptable de dificultad a valores entre 0,3 y 0,8 ^(11,22). La discriminación del ítem fue estimada con la correlación biserial puntual ⁽²³⁾, se consideró los siguientes puntos de corte: pobre discriminación (0 a 0,19), aceptable (0,2 a 0,29), buena (0,3 a 0,39) y excelente ($\geq 0,4$) ^(11,24). La confiabilidad interna de los 28 ítems fue estimada mediante el índice de Kuder-Richadson, adecuado para datos binarios (respuesta correcta o no). Un índice > 0,7 fue considerado aceptable, valores > 0,8 fueron considerados buenos, estos cortes fueron usados en el estudio de validación del TOSLS original ⁽¹¹⁾. Realizamos un análisis de sensibilidad considerando la exclusión de cada uno de los ítems (uno por vez).

Se estimó el índice de dificultad (parámetro *b*) utilizando la teoría de respuesta al ítem (TRI) mediante el modelo de Rasch. Este modelo es adecuado para evaluar preguntas de opción múltiple con mejor respuesta única, en las cuales tenemos una mejor respuesta —puntuada con 1—, mientras que la elección de los distractores puntúa 0. El modelo de Rasch se basa en el supuesto de que la probabilidad de un estudiante para responder correctamente un ítem es una función logística de la diferencia entre la habilidad del estudiante (θ) y la dificultad de la pregunta (*b*) ⁽²⁵⁾.

Los supuestos requeridos para la aplicación de la TRI fueron verificados. El primer supuesto de unidimensionalidad requiere que los ítems midan un único rasgo o constructo subyacente (AC). Para ello se aplicó el análisis de componentes

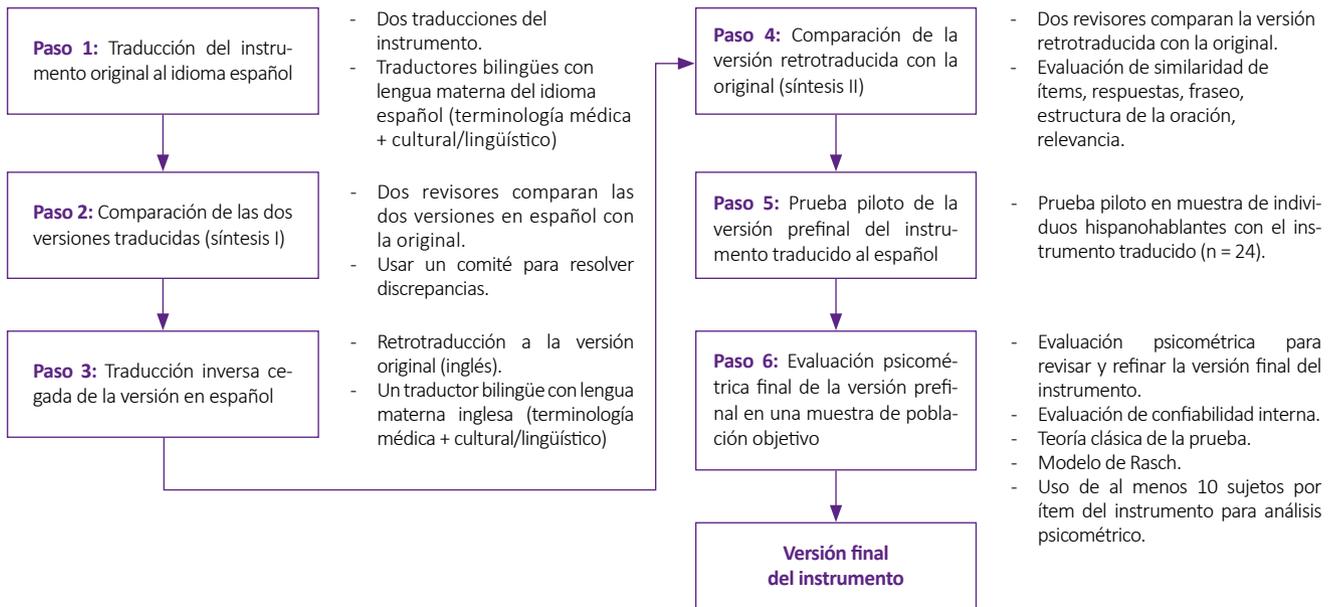


Figura 1. Esquema general de los pasos para la adaptación cultural y validación de un instrumento para medir alfabetismo científico en estudiantes de medicina.

Paso 1: el TOSLS fue traducido al español por dos traductores hispanohablantes, uno con conocimiento en la terminología científica, y un segundo, que, además, era corrector de estilo profesional familiarizado con frases coloquiales y expresiones idiomáticas del español. A ambos traductores se les proporcionó el instrumento original completo, que incluía las indicaciones, las preguntas y sus respectivas opciones, y se les solicitó que generaran dos versiones traducidas de manera independiente.

Paso 2: las dos versiones traducidas y el TOSLS se sometieron a una revisión ítem por ítem por el autor y el traductor con conocimiento en terminología científica. Cualquier discrepancia que surgió fue discutida y resuelta hasta llegar a un consenso, lo que resultó en la creación de una versión preliminar en español.

Paso 3: implicó una traducción inversa, donde la versión preliminar en español del TOSLS se entregó a un traductor cuya lengua materna era el inglés. Este traductor no tenía conocimiento previo del instrumento original.

Paso 4: la versión en inglés resultante de la traducción inversa fue comparada con el instrumento original por el autor (un médico epidemiólogo con experiencia en procesos de enseñanza en investigación científica y en investigación en educación médica) y uno de los traductores del primer paso. Cualquier discrepancia que surgiera se resolvió mediante consenso. En caso de persistir discrepancias, se repitió los pasos 1 a 4 únicamente para los ítems problemáticos (Figura 1). Al culminar el paso 4 se obtuvo una versión pre-final traducida al español, esta versión tuvo equivalencia conceptual, semántica y de contenido con la versión original en inglés.

Paso 5: la versión pre-final traducida fue sometida a una prueba piloto en 24 estudiantes de medicina del séptimo año.

principales (ACP) con el método de rotación *varimax*, el número de factores a extraer fue realizado con el análisis paralelo, una técnica recomendada para identificar el número de factores comunes. Previamente, se verificó que los datos de la muestra tengan adecuación para una estructura factorial, para ello usamos la prueba Kaiser-Meyer-Olkin (KMO) y la prueba de esfericidad de Bartlett. Se consideró un valor de KMO $\geq 0,80$ como satisfactorio ⁽²⁶⁾, y un valor de $p < 0,05$ como significativo para la prueba de Bartlett.

El supuesto de independencia local de los ítems implica que las correlaciones entre ítems fueran generadas por el factor latente medido por el instrumento; y no porque la probabilidad de responder correctamente a un ítem dependa de la respuesta a otro. Para este análisis se estimó el estadístico Q_3 con el método de estimación de máxima verosimilitud conjunta; se

consideró que valores inferiores a 0,5 indicaron independencia local ⁽²⁵⁾. También se estimó el estadístico MADaQ3, los valores cercanos a 0 indicaron la no correlación entre los residuales de los ítems, se consideró valores $\leq 0,10$ para independencia local. Se aplicó el método de Holm para estimar el valor de p de las múltiples pruebas de comparación. El estadístico $Q3$ fue estimado con el método de estimaciones de probabilidad ponderada (WLE, por sus siglas en inglés) ⁽²⁷⁾.

Para evaluar el ajuste de los ítems del instrumento con el modelo de Rasch estimamos los estadísticos basados en residuos cuadráticos ponderados con su varianza. Dos estadísticos, el *infit* (MNSQ *infit*) y el *outfit* (MNSQ *outfit*) fueron calculados. En este análisis se consideró valores entre 0,70 y 1,30 como indicadores de buen ajuste. Valores de *infit* y *outfit* mayores de 1,3 indican ítems inconsistentes

o que no miden el constructo subyacente y valores inferiores a 0,70 indican dependencia de los ítems (redundancia) ⁽²⁵⁾. La medida de confiabilidad de separación de las personas (PSR, *person separation reliability*) se usó para evaluar el grado en que el TOSLS en español discrimina de manera consistente a las personas en función de su habilidad en alfabetismo científico, un valor $> 0,70$ fue aceptable.

Las estimaciones fueron realizadas con el método de máxima verosimilitud marginal (MMLE, por sus siglas en inglés). Los análisis fueron realizados con el programa JAMOMI 2.4.8, en su módulo snowIRT ⁽²⁸⁾.

Aspectos éticos

El estudio fue aprobado por el Comité Institucional de Ética en Investigación de la Universidad de Piura. Aplicamos un consentimiento informado abreviado, el

análisis estadístico fue realizado en una base de datos anonimizada. Los participantes no recibieron estímulos de ningún tipo para participar en el estudio.

RESULTADOS

319 estudiantes fueron invitados, de los cuales 11 no brindaron su consentimiento informado (3 del primer año y 8 del segundo año), además 8 estudiantes del primer año entregaron el cuestionario antes de los primeros 15 minutos. Los

datos de 300 estudiantes ingresaron al análisis: 84 (28,0%) del primer año, 59 (19,7%) del segundo, 31 (10,3%) del tercero, 40 (13,3%) del cuarto, 56 (18,7%) del quinto y 30 (10,0%) del sexto año. Un estudiante no brindó datos de sexo, el 54,8% ($n = 164$) fueron mujeres; entre 295 datos válidos para edad, la media fue 20,4 años (desviación estándar = 2,2) con un mínimo de 17 y máximo de 27 años.

La mediana del puntaje fue 19,0, los percentiles 25 y 75 fueron los puntajes 16,0 y 21,3, respectivamente. El mínimo

fue 5 y el máximo 27. Los puntajes no se adecuaron a una distribución normal (W de Shapiro-Wilk = 0,976, $p < 0,001$).

Índice de dificultad y discriminación de los ítems

El promedio de los índices de dificultad (*idif*) de los 28 ítems fue 0,651. Los ítems con los índice de dificultad superiores a 0,8 fueron cinco: ítem 5 (*idif* = 0,927), ítem 22 (*idif* = 0,913), ítem 9 (*idif* = 0,870), ítem 27 (*idif* = 0,823) y el ítem 23 (*idif* = 0,813). El ítem 12 tuvo un índice de dificultad inferior a 0,3 (*idif* = 0,203) (Figura 2A).

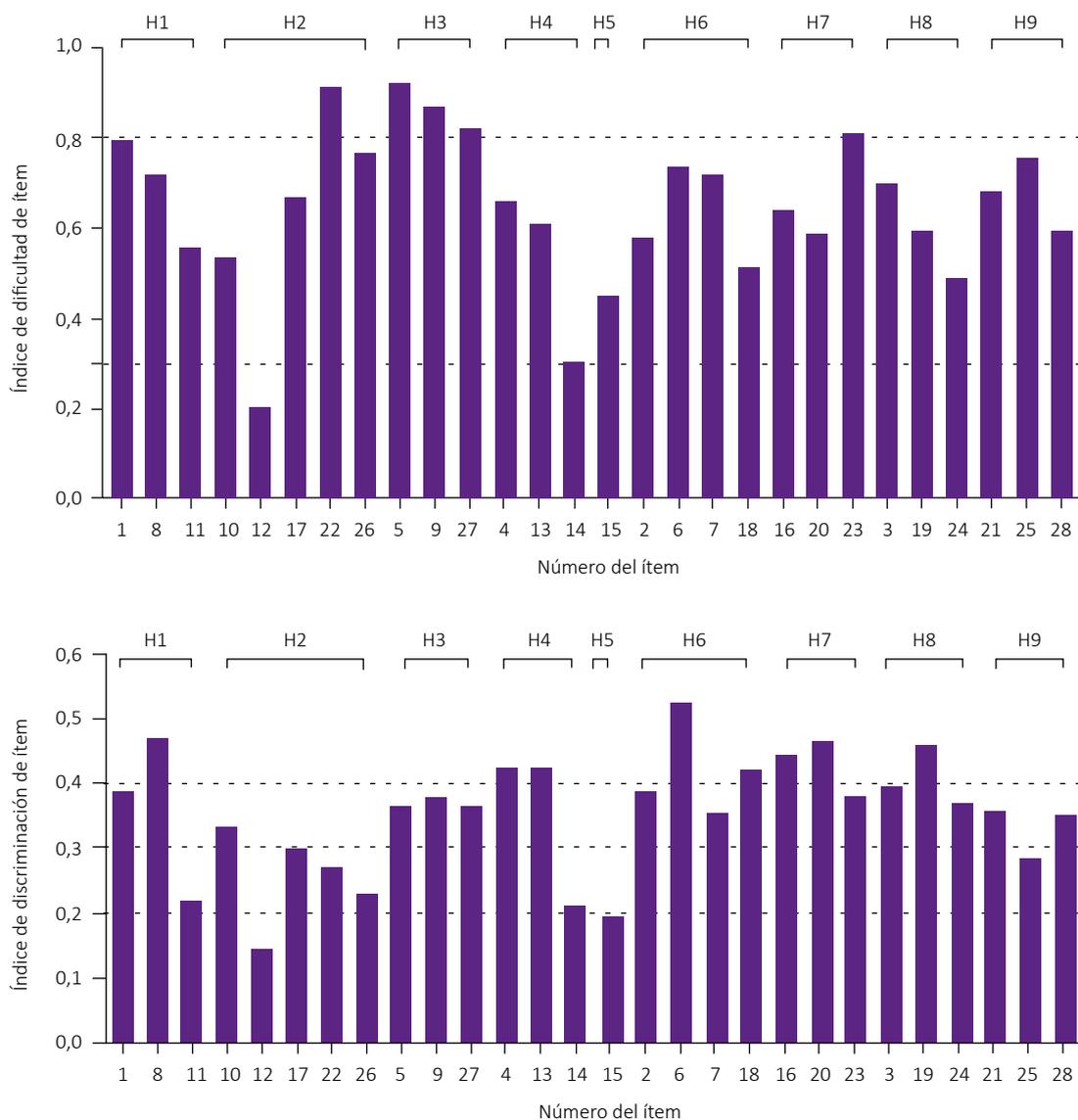


Figura 2. Índice de dificultad y discriminación de los ítems del instrumento para medir habilidades de alfabetismo científico (TOSLS en español).

H1: Identificar un argumento científico válido, H2: Evaluar la validez de la fuente, H3: Evaluar el uso de la información científica, H4: Entender los elementos del diseño de investigación y como ellos impactan los resultados y conclusiones, H5: Crear representaciones gráficas de los datos, H6: Leer e interpretar las representaciones gráficas de los datos, H7: Solucionar problemas usando habilidades cuantitativas, H8: Entender e interpretar las estadísticas básicas, H9: Justificar inferencias, predicciones y conclusiones basadas en datos cuantitativos.

Respecto al índice de discriminación, 14 y 7 ítems tuvieron una buena y excelente capacidad discriminativa entre los estudiantes según su habilidad. Por otro lado, cinco y dos ítems tuvieron capacidades discriminativas regulares y pobres, respectivamente. El promedio de los índices de discriminación de los ítems fue 0,351 [mínimo de 0,145 (ítem 12) - máximo de 0,524 (ítem 6)] (Figura 2B).

Confiabilidad interna de los ítems

La consistencia interna fue 0,742 (aceptable), otros coeficientes de consistencia interna como el alfa de Cronbach y el ω de McDonald tuvieron valores de 0,741 y 0,752, respectivamente. Si se retirase del cuestionario individualmente los ítems 11, 12, 14, 15 y 26, el coeficiente de confiabilidad interna incrementaría entre 0,001 (retiro ítem 26) a 0,005 (retiro ítem 12 o 15) (Material suplementario 1: tabla 2).

Evaluación de la unidimensionalidad

El análisis de componentes principales con el método de rotación de varimax sugirió un único factor, el cual explicó el 14,2% de la varianza observada en los datos. La prueba de esfericidad de Barlett brindó evidencia de adecuación a una estructura factorial ($\chi^2 = 1000$, grados de libertad = 378, $p < 0,001$), además, el índice de Kaiser-Meyer-Olkin corroboró la viabilidad del análisis factorial en la muestra (KMO = 0,739). Cinco ítems tuvieron cargas factoriales menores a 0,2 (ítem 11, 12, 14, 15 y 26) (Material suplementario 1: tabla 3). Se decidió mantener dichos ítems por criterio teórico.

Independencia local de los ítems

La media de los valores absolutos de las correlaciones por pares de los residuos (estadístico MADaQ3) fue 0,054, con un valor de $p < 0,001$ (método de ajuste de Holm). El estadístico Q3 estimado fue 0,437. Se concluyó en que los ítems cumplieron con el supuesto de independencia local.

Estimados de separación de personas y ajuste de ítems al modelo de Rasch

La medida de confiabilidad de las personas fue 0,711, considerado aceptable. Ningún ítem del instrumento tuvo valores de los índices *infit* y *oufit* fuera del rango aceptable (Tabla 1).

Curva de características de los ítems

Entre los ítems que miden las habilidades para entender los métodos de investigación que producen el conocimiento científico, el ítem 12 y el 14 fueron los más difíciles. Los niveles de habilidad del encuestado en la posición $\theta = 3$ generan una probabilidad del 80% de responder correctamente dichos ítems. Por otro lado, el ítem 5 resultó el más fácil (Figura 3). Entre los ítems que miden la habilidad para organizar, analizar e interpretar datos cuantitativos e información científica, el ítem 15 y 23 fueron el más difícil y fácil, respectivamente (Figura 4).

Función de información de la prueba

Se observa que la mayor cantidad de información (confiabilidad) proporcionada por el instrumento se encuentra para los estudiantes con habilidad promedio (logits en posición 0) y cercano a -1 desviación estándar debajo de la media (Material suplementario 1: figura 1A).

Mapa de Wright

Los ítem 12 y 14 fueron los más difíciles del instrumento; los ítems 9, 22 y 5 están ubicados entre los más fáciles (menor que posición -2 en escala logit). 21 de los 28 ítems se encuentra localizados entre -2 a 0 logits de la escala de dificultad. Las medidas de habilidad de los estudiantes oscilan entre -2 logits y +2 logits. Ningún ítem estuvo posicionado en la escala de dificultad por debajo de los estudiantes menos capaces, igualmente, ningún ítem estuvo posicionado en la escala de dificultad por encima de los estudiantes más capaces (Material suplementario 1: figura 1B).

DISCUSIÓN

La versión en español del TOSLS fue validada en una muestra de estudiantes de medicina en Perú. Los resultados sugieren propiedades psicométricas robustas para medir el AC. Estos hallazgos, además, brindan evidencia para que educadores e investigadores puedan decidir sobre su uso conociendo las características psicométricas de sus ítems. El TOSLS en su versión en español del Perú mantiene el número de ítems de la prueba original en inglés⁽¹¹⁾, por lo tanto, permite recoger información sobre nueve habilidades de AC.

Esta versión tuvo medidas psicométricas consistentes con la versión en inglés⁽¹¹⁾. La consistencia interna fue aceptable en ambas versiones, y los valores se asemejaron a los obtenidos en estudiantes de biología en los Estados Unidos. Bajo la teoría clásica de la prueba se estimó un promedio del índice de dificultad de 0,65, similar a lo encontrado en el estudio original que obtuvo valores entre 0,59 y 0,68. Además, el índice de discriminación se situó en 0,351, superando el valor de 0,27 informado en la validación de la versión original entre estudiantes de biología. El TOSLS en español de Perú demostró una buena capacidad de discriminación.

El análisis de Rasch aplicado al TOSLS en español brinda información nueva sobre sus ítems. En primer lugar, corroboramos los supuestos requeridos para proceder con este modelo. Tanto la versión en español como la versión en inglés tienen una estructura unidimensional. La independencia local de los ítems se cumplió, lo que implicó que las diferencias entre los ítems se relacionaron con las habilidades en AC de los estudiantes, y la probabilidad de responder correctamente a un ítem fue independiente de otros. En general, los ítems del TOSLS versión en español se ajustaron al modelo de Rasch. Ningún ítem tuvo características psicométricas que evidenciaran falta de consistencia o que no midan el constructo subyacente.

El TOSLS en español presenta ítems con diferentes niveles de dificultad. El ítem 12 resultó el de mayor dificultad, este ítem, y junto con otros cuatro ítems, evalúa la habilidad de los participantes para evaluar la validez de una fuente bibliográfica. El ítem 14 fue el segundo más difícil, este mide la habilidad para entender los elementos del diseño de investigación y como estos impactan los resultados y conclusiones. Por otro lado, las preguntas más fáciles fueron los ítems 9, 22 y 5 (en orden ascendente). Los ítems 5 y 9 son dos de los tres ítems que miden la habilidad de evaluar el uso de la información científica; el ítem 22 junto al ítem 12 miden la validez de una fuente bibliográfica.

Otros estudios han estimado el nivel de dificultad con la teoría clásica de la prueba, con resultados diversos. En el estudio de Gormally *et al.*⁽¹¹⁾ el ítem 14 fue el más

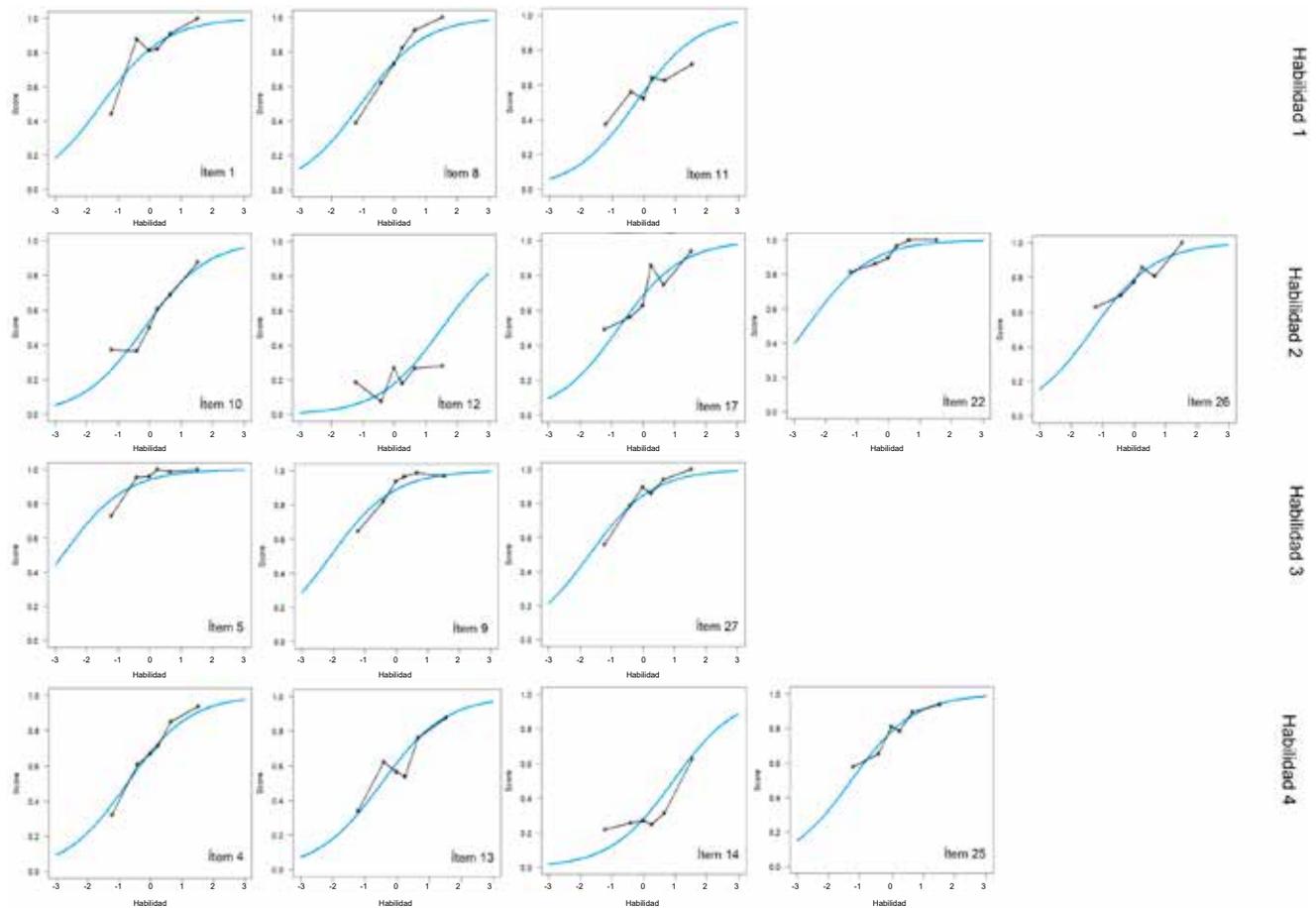


Figura 3. Curvas características de los ítems que miden las habilidades para entender los métodos de investigación que producen el conocimiento científico.

difícil para los estudiantes de biología, seguido del ítem 17 y 15. Entre estudiantes del primer año de medicina de la India, se encontró que menos del 1% respondió correctamente los cuatro ítems que miden la habilidad de leer e interpretar las representaciones gráficas de los datos, además el 1,52% de estudiantes respondió bien los ítems de la habilidad para evaluar la validez de la fuente ⁽¹⁹⁾. En profesores de licenciatura en Brasil, los ítems 6, 10, 12, 16, 19 y 22 fueron los más difíciles, con menos del 25% de participantes que acertaron; mientras que las más fáciles fueron los ítems 1 y 27 con aproximadamente 90% de acierto ⁽¹⁸⁾.

La función de información de la prueba muestra que el TOSLS en español fue más preciso en determinados niveles de habilidad de AC, específicamente en aquellos con un nivel promedio de habilidad correspondiente a una desviación estándar por debajo de la media. Esta característica sugiere que esta versión del TOSLS podría ser particularmente valiosa

como una herramienta de certificación ⁽²¹⁾, es decir, que podría ser útil para determinar si los estudiantes cumplen con los requisitos necesarios para aprobar cursos destinados al desarrollo de habilidades en AC.

Cuando se comparan la distribución de las habilidades en AC de los estudiantes y la distribución de las dificultades de los ítems en una misma escala, se observa atributos de una evaluación equilibrada. El TOSLS en español del Perú mide la habilidad del estudiante y ésta tiene correspondencia con los niveles de dificultad de los ítems, es decir, los estudiantes más hábiles en AC tienden a responder a las preguntas más desafiantes. Aunque no hay una simetría perfecta en la distribución, se confirma que el nivel de habilidad de los estudiantes no es superado por la dificultad de los ítems en ningún extremo de la escala.

El presente estudio tuvo limitaciones; entre los alumnos del primer y segundo

año hubo una mayor tasa de no respuesta, sin embargo, no afectó el tamaño muestral requerido. La selección muestral fue no probabilística lo cual podría afectar la heterogeneidad en el rasgo latente de AC, sin embargo, el estudio incluyó estudiantes del primer al sexto año para asegurar la variabilidad requerida en AC. Por otro lado, el análisis basado en la teoría de respuesta al ítem dependen menos de la muestra, lo cual reduce la afectación en la generalización de estos hallazgos. La traducción inversa fue realizada por un solo traductor, sin embargo, en la revisión de la síntesis se verificó la consistencia de esta única traducción con la versión original. En Perú, la carrera de medicina dura siete años, el séptimo grupo participó en el piloto, más no en la validación psicométrica; sin embargo, la inclusión de estudiantes del quinto y sexto año aseguró individuos con mayores niveles de habilidad en AC.

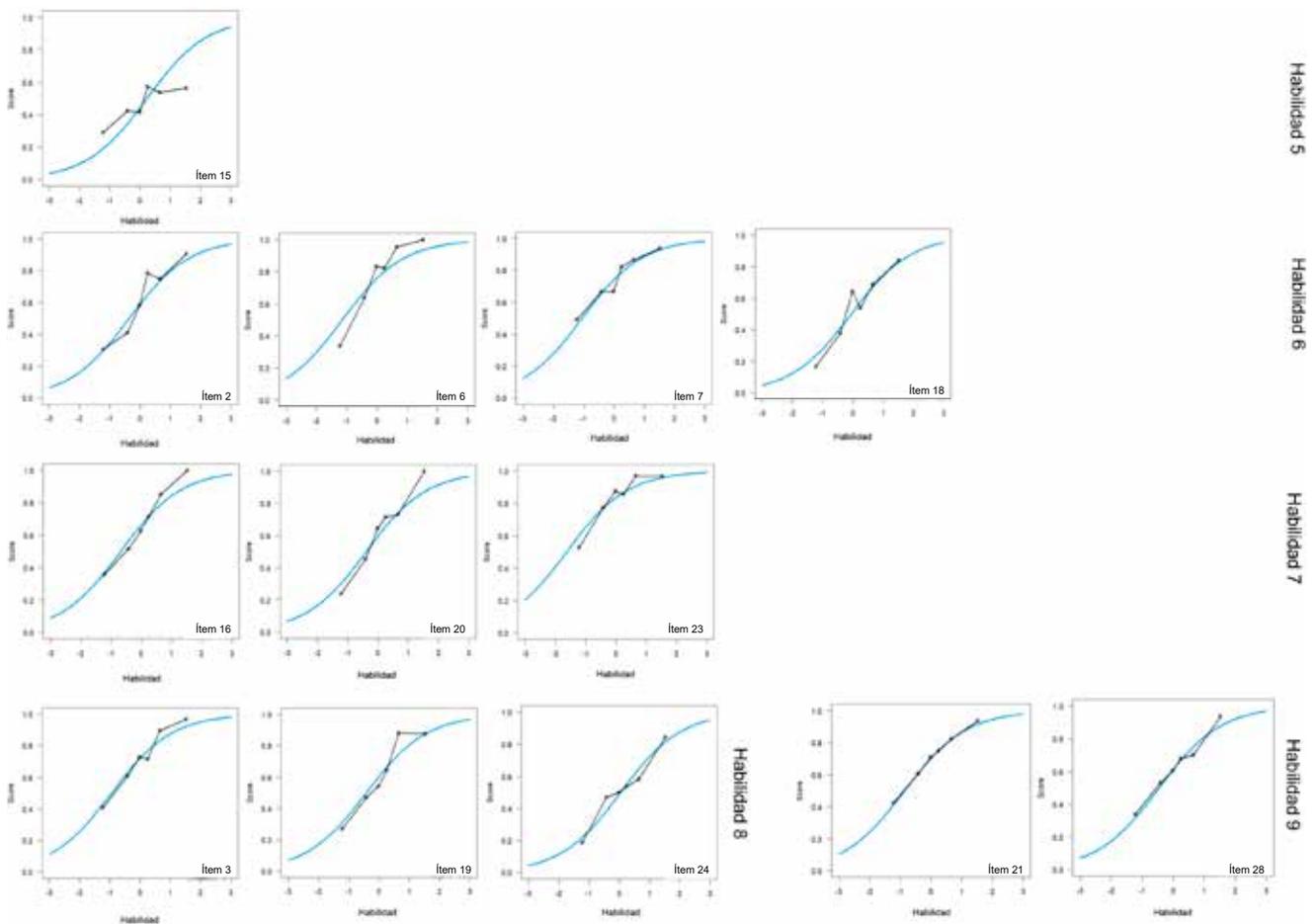


Figura 4. Curvas características de los ítems que miden la habilidad para organizar, analizar e interpretar datos cuantitativos e información científica.

Los resultados deben ser interpretado con cautela, en primer lugar, la adaptación y validación fue realizada al español hablado en Perú, por ello, los resultados podrían no ser generalizados a otras poblaciones hispanohablantes. Segundo, si bien esta validación fue realizada en estudiantes de medicina y el análisis realizado brinda información de la calidad intrínseca de los ítems del TOSLS en español, independientemente de la habilidad en AC de los estudiantes; es recomendable evaluar las propiedades psicométricas del instrumento en otras carreras universitarias.

Este hallazgo es relevante tanto para la investigación en educación médica como para la mejora de los programas de enseñanza en las facultades de medicina. Esta versión validada del TOSLS está disponible libremente para su utilización en investigaciones que busquen evaluar el AC en estudiantes de medicina hispa-

nohablantes (Material Suplementario 2). Además, puede emplearse como herramienta para evaluar el resultado generado por intervenciones diseñadas para mejorar el nivel de AC. Desde una perspectiva educativa, contar con una herramienta validada permitirá la evaluación de estas habilidades en el primer año de estudio, lo que facilitará la medición de los cambios generados por cursos orientados a formar o fortalecer competencias en AC al inicio de la carrera médica

En conclusión, presentamos la versión al español del Perú de una «Prueba de Habilidades en Alfabetismo Científico» (TOSLS en español), la cual mantuvo el número original de ítems de su versión en inglés. Este instrumento por su adecuación al modelo de Rasch presenta cualidades psicométricas que nos permiten tener una aproximación cercana al nivel de AC en estudiantes de medicina de todos los años de la carrera.

AGRADECIMIENTOS

A César Gutiérrez por su apoyo en la planificación y administración de la prueba.

REFERENCIAS BIBLIOGRÁFICAS

1. OECD. PISA 2018 Science Framework [Internet]. 2019. Disponible en: <https://www.oecd-ilibrary.org/content/component/f30da688-en>
2. Committee on Science Literacy and Public Perception of Science, Board on Science Education, Division of Behavioral and Social Sciences and Education, National Academies of Sciences, Engineering, and Medicine. Science Literacy: Concepts, Contexts, and Consequences [Internet]. Snow CE, Dibner KA, editores. Washington, D.C.: National Academies Press; 2016 [citado el 11 de mayo de 2023]. DOI:10.17226/23595
3. Impey C, Buxner S, Antonellis J, Johnson E, King C. A Twenty-Year Survey of Science Literacy Among College Undergraduates. *J Coll Sci Teach* [Internet]. 2011 [citado el 11 de mayo de 2023]; 40(4): 31-37. Disponible en: <https://www.depauw.edu/files/resources/impey2011.pdf>
4. Shaffer JF, Ferguson J, Denaro K. Use of the Test of Scientific Literacy Skills Reveals That Fundamental Literacy Is an Important Contributor to Scientific

- Literacy. *CBE—Life Sci Educ.* 2019;18(3):ar31. DOI:10.1187/cbe.18-12-0238
5. Porter JA, Wolbach KC, Purzycki CB, Bowman LA, Agbada E, Mostrom AM. Integration of Information and Scientific Literacy: Promoting Literacy in Undergraduates. *CBE—Life Sci Educ.* 2010;9(4):536–42. DOI:10.1187/cbe.10-01-0006
 6. Vision and change in undergraduate biology education. A call to action. A final report of a National Conference organized by the American Association for the Advancement of Science [Internet]. Washington, D.C; 2009. Disponible en: <https://www.aps.org/programs/education/undergrad/upload/Revised-Vision-and-Change-Final-Report.pdf>
 7. Auerbach AJ, Schussler EE. Curriculum Alignment with *Vision and Change* Improves Student Scientific Literacy. *CBE—Life Sci Educ.* 2017;16(2):ar29. DOI:10.1187/cbe.16-04-0160
 8. Kelp NC, McCartney M, Sarvary MA, Shaffer JF, Wolyniak MJ. Developing Science Literacy in Students and Society: Theory, Research, and Practice. *J Microbiol Biol Educ.* 2023; e00058-23. DOI:10.1128/jmbe.00058-23
 9. Coppi M, Fialho I, Cid M. Scientific literacy assessment instruments: a systematic literature review. *Educ Em Rev.* 2023;39: e37523. DOI:10.1590/0102-4698237523-t
 10. Laugksch RC, Spargo PE. Construction of a paper-and-pencil Test of Basic Scientific Literacy based on selected literacy goals recommended by the American Association for the Advancement of Science. *Public Underst Sci.* 1996;5(4):331–59. DOI:10.1088/0963-6625/5/4/003
 11. Gormally C, Brickman P, Lutz M. Developing a Test of Scientific Literacy Skills (TOSLS): Measuring Undergraduates' Evaluation of Scientific Information and Arguments. *CBE—Life Sci Educ.* 2012;11(4):364–77. DOI:10.1187/cbe.12-03-0026
 12. Fives H, Huebner W, Birnbaum AS, Nicolich M. Developing a measure of scientific literacy for middle school students: Developing a measure of scientific literacy. *Sci Educ.* 2014;98(4):549–80. DOI:10.1002/sce.21115
 13. Mun K, Lee H, Kim S-W, Choi K, Choi S-Y, Krajcik JS. Cross-cultural comparison of perceptions on the global scientific literacy with Australian, Chinese, and Korean middle school students. *Int J Sci Math Educ.* 2015;13(S2):437–65. DOI:10.1007/s10763-013-9492-y
 14. Segarra VA, Hughes NM, Ackerman KM, Grider MH, Lyda T, Vigueira PA. Student performance on the Test of Scientific Literacy Skills (TOSLS) does not change with assignment of a low-stakes grade. *BMC Res Notes.* 2018;11(1):422. DOI:10.1186/s13104-018-3545-9
 15. Cartwright NM, Liddle DM, Arceneaux B, Newton G, Monk JM. Assessing scientific literacy skill perceptions and practical capabilities in fourth year undergraduate biological science students. *Int J High Educ.* 2020;9(6):64. DOI:10.5430/ijhe.v9n6p64
 16. Waldo JT. Application of the Test of Scientific Literacy Skills in the Assessment of a General Education Natural Science Program. *J Gen Educ.* 2014;63(1):1–14. DOI:10.5325/jgeneeduc.63.1.0001
 17. Santiago D, Nunes A, Alves L. Letramento científico e crenças CTSA em estudantes de pedagogia. *REPPE: Revista Do Programa de Pós-Graduação Em Ensino* [Internet]. 2020 [citado el 21 de abril de 2023]; 4(2):210–36. Disponible en: <https://seer.uenp.edu.br/index.php/reppe/article/view/971>
 18. Gomes ASA, Almeida ACPCD. Letramento científico e consciência metacognitiva de grupos de professores em formação inicial e continuada: um estudo exploratório. *Amaz Rev Educ Em Ciênc E Matemáticas.* 2016;12(24):53. DOI:10.18542/amazrecm.v12i24.3442
 19. Mohan L, Singh Y, Kathrotia R, Cariappa M, Khera A, Ghosh S. Scientific literacy and the medical student: A cross-sectional study. *Natl Med J India.* 2020;33(1):35. DOI:10.4103/0970-258X.308242
 20. Sousa VD, Rojjanasriat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline: Validation of instruments or scales. *J Eval Clin Pract.* 2011;17(2):268–74. DOI:10.1111/j.1365-2753.2010.01434.x
 21. De Champlain AF. A primer on classical test theory and item response theory for assessments in medical education: Classical test theory and item response theory. *Med Educ.* 2010;44(1):109–17. DOI:10.1111/j.1365-2923.2009.03425.x
 22. Johari J, Sahari J, Wahab DA, Abdullah S, Abdullah S, Omar MZ, et al. Difficulty Index of Examinations and Their Relation to the Achievement of Programme Outcomes. *Procedia - Soc Behav Sci.* 2011;18:71-80. DOI: 10.1016/j.sbspro.2011.05.011
 23. Elaboración de preguntas para evaluaciones escritas en el área de ciencias de la salud. National Board of Medical Examiners [Internet]. 2022. Disponible en: https://www.nbme.org/sites/default/files/2022-10/NBME_Item-Writing_Guide_Spanish.pdf
 24. Rao C, Kishan Prasad H, Sajitha K, Permi H, Shetty J. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. *Int J Educ Psychol Res.* 2016;0(0):0. DOI: 10.4103/2395-2296.186515
 25. Tavakol M, Dennick R. Psychometric evaluation of a knowledge based examination using Rasch analysis: An illustrative guide: AMEE Guide No. 72. *Med Teach.* 2013;35(1):e838–48. DOI: 10.3109/0142159X.2012.737488
 26. Lloret-Segura S, Ferreres-Traver A, Hernández-Baeza A, Tomás-Marco I. El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada. *An Psicol.* 2014;30(3):1151–69. DOI: 10.6018/analeps.30.3.199361
 27. Robitzsch A, Kiefer T, Wu M. TAM: Test Analysis Modules. R package version 4.1-4. [Internet]. 2022. Disponible en: <https://CRAN.R-project.org/package=TAM>.
 28. Seol H. snowIRT: Item Response Theory for jamovi. (Version 4.9.5) Jamovi module. [Internet]. 2023. Disponible en: <http://github.com/hyunsooseol/snowIRT>.