

La búsqueda de las evidencias de validez en evaluación educativa: ¿Podemos tender puentes entre teoría y práctica?

The search for validity evidence in educational evaluation: Can we build bridges between theory and practice?

“La investigación sobre validez utiliza teoría, datos y lógica para argumentar a favor o en contra de interpretaciones específicas de los resultados”.

STEVE DOWNING¹

El tema de la validez en evaluación educativa es uno de los más apasionantes y complejos laberintos conceptuales y metodológicos que existen en educación. Cuando hablamos de exámenes de ingreso, graduación o certificación, de evaluaciones sumativas de alto impacto, así como las consecuencias de este tipo de evaluaciones, la validez siempre está en la discusión de forma explícita o tácita¹. Hace varios años dedicamos una editorial al tema con la siguiente pregunta en el título: “¿por qué seguimos usando ideas obsoletas?”². Aquí estamos 8 años después, momento en que podemos hacer la misma pregunta con mayor intensidad, pero ahora en un contexto global mucho más complicado por las consecuencias de la pandemia, el creciente desdén y cuestionamiento de los exámenes sumativos de alto impacto en muchos países, la cada vez mayor avalancha de desinformación y noticias falsas, el escepticismo social hacia la ciencia y la academia,

así como el disruptivo escenario de la inteligencia artificial generativa en todos los aspectos de la educación superior, incluyendo la evaluación del y para el aprendizaje.

El concepto de validez en evaluación educativa continúa siendo fundamental en este campo, la investigación sobre el tema ha seguido creciendo y haciéndose cada vez más sofisticada, generando múltiples perspectivas y controversias que no han agotado el tema, sino que incluso lo han hecho más difícil de entender y de aplicar en la práctica. Debemos recordar que el concepto actual de validez ha superado al tradicional en muchos aspectos teóricos y prácticos, promoviendo un modelo mental compartido del tema en la comunidad académica que se dedica a la evaluación educativa³. Las propuestas de Samuel Messick y Michael Kane han servido para nutrir el debate y promover algo de orden sobre el tema, modelos que en el mundo de la educación en profesiones de la salud han sido incorporados y resumidos por diversos autores, entre ellos Steve Downing y David Cook⁴. En palabras de Kane: “Validar una interpretación propuesta o un uso de los puntajes de exámenes es evaluar las afirmaciones

Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

basadas en esas puntuaciones. La mezcla específica de evidencia necesaria para la validación depende de las inferencias que se hagan y los supuestos que se utilicen”⁵.

La edición más reciente de los *Estándares para Pruebas Educativas y Psicológicas* de la AERA, APA y NCME, afirma: “La validez se refiere al grado en que la evidencia y la teoría respaldan las interpretaciones de los puntajes de una prueba para usos propuestos de las pruebas”³. Por ello Downing y Cook argumentan que el concepto actual de validez se refiere a la interpretación de las puntuaciones obtenidas en los exámenes, así como a las inferencias que es aceptable hacer de dichos resultados de acuerdo al diseño y propósito de la prueba^{1,4}. No es la prueba la que es válida o no, sino las inferencias y afirmaciones que se hacen de los resultados, utilizando teorías y hallazgos de evidencia investigacional para darle significado a dichos resultados. De esta manera la validez contribuye a darle significado a las evaluaciones, en virtud de que las mediciones en ciencias sociales tienen poco significado intrínseco, deben interpretarse a la luz del contexto y de la metodología usada para diseñar, aplicar e interpretar las pruebas. En el modelo vigente, la validez no es una característica intrínseca de los instrumentos de evaluación (de ahí que es incorrecto hablar de “el instrumento es válido”, o “se validó el instrumento”), sino que es un concepto holístico que se refiere a la validez de constructo como un todo integrado. Dicha validez de constructo se alimenta de diferentes fuentes: contenido, proceso de respuesta, estructura interna, relación con otras variables, y consecuencias^{1,3,4}.

Todo ello ha sido ampliamente analizado por diversos autores, conformando un esquema aceptado por la comunidad académica de evaluación educativa, y nutriéndose de argumentos elaborados por diferentes autores. Cook y colaboradores plantean el uso del modelo de Kane para proponer una guía práctica del proceso de validación, que incluye articular premisas y afirmaciones con la decisión propuesta por el examen, lo que llaman “argumento de interpretación/uso”, ya que utiliza cuatro inferencias en el argumento de validez: puntuación, generalización, extrapolación e implicaciones. Esta propuesta está ampliamente descrita con ejemplos relevantes

para la comunidad médica en el trabajo de Cook y colaboradores⁴.

Recientemente, después de todo lo que ocurrió en evaluación educativa como resultado de la pandemia por Covid-19, Russell propuso una terminología para las diferentes etapas de la validación, que pretende proporcionar un marco interpretativo menos complejo que el de los modelos abstractos de Messick y Kane⁶. Este modelo reciente fue analizado por nuestro grupo con motivo del trabajo de doctorado de Blanca Carrillo, ensayo publicado en la literatura internacional, invitamos a nuestros lectores a consultarlo⁷. La citada propuesta del concepto de validez se usó como marco conceptual en el trabajo de la Dra. Carrillo, publicado en este número, en el que se analizaron las fuentes de evidencia de validez del proceso de admisión a la Facultad de Medicina de la Universidad Autónoma de San Luis Potosí⁸. Como podemos ver en este trabajo, aún tenemos mucho camino por recorrer en el complejo y sinuoso sendero de la validez en nuestros procesos de evaluación sumativa, como son los exámenes de ingreso a las escuelas de medicina.

Por si el tema no fuera lo suficiente complejo y denso, la aparición en el escenario educativo de la inteligencia artificial generativa ha venido a complicar de forma profunda y todavía incierta, lo que podemos y debemos hacer con los resultados de la aplicación de instrumentos de evaluación, como lo empiezan a describir varios autores⁹. ¡En verdad que vivimos tiempos interesantes en educación superior!

En este número de la revista, además del trabajo de la Dra. Carrillo, contamos con artículos originales sobre los siguientes temas: carga mental en simulación de baja fidelidad, maltrato en las residencias médicas, gestión de aprendizaje autorregulado, grado de cumplimiento de los derechos de médicos residentes, características psicométricas de la escala de depresión, ansiedad y estrés en estudiantes universitarios, gamificación en anatomía, simulación con enfoque de género, percepción de estudiantes de enfermería sobre ayudantes de cátedra. Además, tenemos un artículo de revisión sobre el aprendizaje en el sitio de trabajo, y uno metodológico sobre simulación clínica para desarrollo de razonamiento clínico en estudiantes de kinesiología.

El concepto de validez continúa siendo un extraordinario reto para la comunidad de educadores en profesiones de la salud, ya que se trata de una palabra con una multiplicidad de significados y perspectivas epistemológicas¹⁰. Por ello es indispensable continuar estudiando el tema y realizando trabajos

de investigación originales en nuestro contexto, para avanzar en la aplicación de los mecanismos de evaluación de, para y como aprendizaje, de una forma más profesional y académica, para beneficio de la comunidad docente, el estudiantado, y, en última instancia, de los pacientes a quienes servimos.



Melchor Sánchez Mendiola

EDITOR

Facultad de Medicina, UNAM

REFERENCIAS

1. Downing SM. Validity: On the meaningful interpretation of assessment data. *Med Educ.* 2003;37(9):830-7.
2. Sánchez-Mendiola M. “Mi instrumento es más válido que el tuyo”: ¿Por qué seguimos usando ideas obsoletas? *Inv Ed Med.* 2016;5(19):133-5.
3. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing.* 6th ed. American Educational Research Association. Washington, D. C.: American Educational Research Association, American Psychological Association & National Council on Measurement in Education; 2014. 243 p.
4. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: A practical guide to Kane’s framework. *Med Educ.* 2015;49(6):560-75.
5. Kane MT. Validating the Interpretations and Uses of Test Scores. *J Educ Meas.* 2013;50(1):1-73.
6. Russell M. Clarifying the Terminology of Validity and the Investigative Stages of Validation. *Educational Measurement: Issues and Practice.* 2022;41(2):25-35.
7. Carrillo-Avalos BA, Leenen I, Trejo-Mejía JA, Sánchez-Mendiola M. Bridging Validity Frameworks in Assessment: Beyond Traditional Approaches in Health Professions Education. *Teaching and Learning in Medicine.* 2023; 1-10. Advance online publication. <https://doi.org/10.1080/10401334.2023.2293871>
8. Carrillo-Avalos BA, Leenen I, Trejo-Mejía A, Sánchez-Mendiola M. Evidencias de validez del proceso de admisión a una escuela de medicina en México. *Inv Ed Med.* 2024;13(50):37-55.
9. Azzam T. Artificial intelligence and validity. *New Directions for Evaluation.* 2023;(178-179):85-95. <https://doi.org/10.1002/ev.20565>
10. St-Onge C, Young M, Eva KW, Hodges B. Validity: one word with a plurality of meanings. *Advances in Health Sciences Education.* 2017;22(4):853-867.