

## Validez y confiabilidad de los procesos evaluativos: El docente retomando las riendas de la evaluación

*Ante un contexto laboral más exigente y complejo en ciencias de la salud, por un paradigma que transitó de la óptica biomédica a la biopsicosocial, y con un escenario universitario en expansión desregulada que ha derivado en una población crecientemente heterogénea de estudiantes, docentes e incluso instituciones, se hace evidente la necesidad de crear instrumentos de evaluación que certifiquen qué tan preparados están los profesionales para enfrentar el mundo laboral. También se hace evidente la necesidad de mejorar las evaluaciones (sumativas y formativas) durante todo el transcurso de la carrera. Lo que no es tan evidente es cómo asegurar la calidad de estas evaluaciones y cómo usar correctamente sus resultados.*

*El énfasis en las evaluaciones sumativas ha concentrado la atención de todos en las calificaciones o los puntajes que entregan, convirtiendo al número final en una entidad intrínsecamente valiosa a la que se le atribuyen altísimos niveles de exactitud y poder de decisión, como si fuese derivada de un ejercicio matemático puro. Parecemos olvidar cuando tenemos un alumno ante nosotros, que el objetivo de toda evaluación es recoger información sobre los avances que él ha logrado, y que tanto la selección de los contenidos que se evaluaron como el diseño del instrumento que se ocupó, descansan en un proceso teórico y razonado, pero ante todo, profundamente humano; y por tanto, toda decisión que tomemos a partir de estos puntajes descansan en niveles de error, muchos de los cuales están más allá de lo que las herramientas estadísticas –cuando menos las actuales– pueden calcular.*

*Cualquier puntaje que obtengamos proviene de un proceso de medición, y éste no es más que un procedimiento que permite asignar valores numéricos a las unidades (en este caso, personas) para reflejar el nivel en el que presentan un fenómeno determinado (p.e. aprendizajes, competencias, intereses, etc.).... Ahora, existen dos propiedades que permiten analizar la calidad de estos procesos de medición: la confiabilidad y la validez. Lamentablemente, el desconocimiento o la malinterpretación de estas propiedades es bastante extendida en educación –no sólo en carreras de la salud–.*

*La primera se refiere al grado de precisión de una medición, entendiendo ésta como aquella parte del puntaje obtenido que no se debe a errores aleatorios, que son los que pueden derivar de los contenidos que el docente decidió incluir en la medición (de todos los contenidos posibles a incluir), las personas que calificaron las respuestas (de todos los calificadores posibles) o el momento en que se decidió evaluar (de todos los momentos posibles). Así, cada elección que se toma en el proceso puede ser una fuente de error en la medición.*

*La segunda, la validez, desde los años '80 ya no se entiende como que un instrumento mida lo que pretende medir (la acepción que usualmente se le asigna), sino que ahora se entiende como el grado de apoyo teórico y empírico para la interpretación y el uso que hacemos de las puntuaciones que entrega<sup>1</sup>.*

*De esta forma, la validez no es una propiedad del instrumento sino de las conclusiones que sacamos de él. Por ejemplo, la pregunta no es si un examen con preguntas de opción múltiple mide o no conocimiento sobre anatomía, sino si podemos decir que un alumno que saca un 3,9 en dicho examen no ha alcanzado los aprendizajes suficientes en la asignatura o, yendo más allá, si podemos decidir que debe reprobado ésta. Responder si mide anatomía parece más sencillo, y es el paso necesario para responder lo demás: sin embargo, las últimas dos interrogantes implican otras reflexiones teóricas, empíricas, prácticas e incluso éticas.*

*Un ejemplo paradigmático de malas interpretaciones –por tanto, interpretaciones no válidas– podemos encontrarlo en la Prueba de Selección Universitaria de Chile, la PSU. Este instrumento fue diseñado para jerarquizar a los estudiantes de acuerdo a los contenidos aprendidos durante la enseñanza media, para así poder elegir a los más aptos de cada cohorte (los que han aprendido más, y por tanto tienen más respuestas correctas). Sin embargo, la comunidad en general, la prensa e incluso algunas autoridades han insistido en utilizarla como un baremo para evaluar la calidad de la educación. De esta forma, insisten en comparar los resultados PSU de distintos tipos de establecimientos, ignorando que esta prueba no cubre todos los aprendizajes relevantes que se esperan en enseñanza media (porque, como se dijo, sólo aborda contenidos), o se satisfacen al saber que más alumnos alcanzan los puntajes nacionales teniendo todas las respuestas correctas, ignorando que esto ilustra un efecto techo, esto es, una progresiva incapacidad de la PSU para diferenciar y jerarquizar a los sujetos de niveles más altos, siendo que*

1 Martínez M R, Hernández M J, Hernández V. Psicometría 2006. Madrid: Síntesis.

2 Koljatic M, Silva M. Validación de la PSU: Comentarios al «Estudio acerca de la validez predictiva de los factores de selección a las universidades del Consejo de Rectores». Estudios públicos 2006; 104: 331-346.

3 Pérez C, Ortiz L, Parra P. Prueba de selección universitaria, rendimiento en enseñanza media y variables cognitivo-actitudinales de alumnos de medicina. Revista de Educación en Ciencias de la Salud 2011; 8(2): 120-127.

4 www.eunacom.cl

*jerarquizar es justamente el objetivo de la prueba.*

*Allende los problemas que la PSU tiene para cumplir con los objetivos que sí se propone (tema ya identificado en otros estudios<sup>2,3</sup>), es cada vez más extendida la mala comprensión y el mal uso de los puntajes de esta prueba, pese a los intentos del mismo DEMRE por evitarlo.*

*Caso similar ocurre con el EUNACOM, un instrumento que supone una mejora del antiguo Examen Médico Nacional, al sumar una evaluación de conocimientos prácticos al tradicional examen teórico. En la actualidad esta prueba se utiliza como filtro para ser contratado por organismos públicos, prestar atención por FONASA y postular a becas<sup>4</sup>, lo que es comprensible si se asume que es necesario un nivel mínimo de conocimientos para realizar estas actividades.*

*Más allá de discutir si en su estado actual el EUNACOM permite certificar que un médico está habilitado para estas actividades, y por tanto cuestionarse si es válido usarlo para estos fines, es evidente que la comunidad ha empezado a realizar interpretaciones inválidas sobre el mismo.*

*A modo de ejemplo, la prensa frecuentemente ha utilizado los resultados de este examen como una herramienta para jerarquizar a las universidades, como si fuese una medida de la calidad de la formación médica. Y este mismo error conceptual ha sido aprovechado por algunas casas de estudio para promocionar sus propios programas, al obtener puntajes destacados entre las Escuelas de Medicina nacionales.*

*¿Por qué lo anterior constituiría una interpretación inválida? Porque implicaría asumir que los conocimientos son los únicos aprendizajes que debe alcanzar un médico para egresar de la carrera o, inclusive, para ejercer adecuadamente su rol... Y esto implicaría negar directamente los modelos curriculares de la mayoría, si no es que todas, las Escuelas de Medicina chilenas que hoy trabajan desde un modelo por competencias (u orientado por éstas) y declaran en sus Currícula que desarrollar actitudes y habilidades profesionalmente relevantes, es tan necesario como memorizar conceptos.*

*Lamentablemente, no sólo la prensa y los publicistas caen en este juego. En ocasiones, muchos docentes y alumnos también lo hacen, y enfocan sus esfuerzos en aprender a descifrar los misterios de las preguntas de opción múltiple (reminiscencias de su adiestramiento en los preuniversitarios), como si el país requiriera mejores puntajes de EUNACOM y no mejores médicos.*

*Los ejemplos anteriores serán un error cada vez más común en la medida que busquemos ciegamente certificaciones y éstas sean administradas por autoridades que desconocen los fundamentos y las limitaciones de los procesos evaluativos. Para frenar esto, es urgente partir por los educadores, quienes en lugar de divinizar las puntuaciones de las evaluaciones deben asumir un rol crítico ante la confiabilidad y la validez de la forma en que las interpretan. Una lectura errónea de éstas puede llevar equivocadamente a reprobar a un alumno que no lo merece, titular a un profesional con serias deficiencias u orientar un proceso formativo hacia los test estandarizados, jibarizándolo.*

*Al contrario, por difícil que parezca, al evaluar, cada docente debe sumergirse en un escenario de consideraciones éticas y técnicas, donde él debe asumir la responsabilidad y el protagonismo en el proceso de evaluación que equivocadamente le hemos dejado a los números.*

**Mag. Cristhian Pérez V.**

*Prof. Asistente, Depto. Educación Médica  
Facultad de Medicina, Universidad de Concepción*