

RUEDA BELTRÁN, M. (coord.) (2011). *¿Evaluar para controlar o para mejorar? Valoración del desempeño docente en las universidades*. México: UNAM, Instituto de Investigaciones sobre la Universidad y la Educación (IISUE), 238 páginas.

Mario Rueda Beltrán, coordinador de la obra, es doctor en Ciencias de la Educación por la Universidad de París VIII (Francia) e Investigador Titular del Instituto de Investigaciones sobre la Universidad y la Educación (IISUE) de la UNAM, así como uno de los precursores y coordinador de la Red Iberoamericana de Investigadores sobre Evaluación de la Docencia (RIIED).

El conjunto de trabajos que se presenta en este libro es parte de las actividades llevadas a cabo por la Red de Investigadores en Evaluación de la Docencia de México (RIED) —que integra a académicos de diversas universidades— dentro del proyecto de investigación *Diagnóstico de la evaluación de la docencia en las universidades mexicanas y prueba de un modelo en distintos contextos institucionales*. El objetivo del mismo fue llevar a cabo un diagnóstico acerca del modo en que se realiza la evaluación del desempeño docente en las universidades, así como aportar alternativas en la evaluación que valorizaran las actividades docentes en un marco de mejora permanente.

Las dificultades para desarrollar un trabajo de estas características son muy variadas, comenzando por la enorme cantidad de instituciones universitarias y su heterogeneidad. El sistema de educación superior en México incluye 2.017 instituciones universitarias (689 públicas

y 1.408 privadas) e integra —según informa la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES, 2006)— universidades de perfil muy diverso: tecnológicas, políticas, pedagógicas e interculturales; así como otros tipos de instituciones (institutos tecnológicos, centros de investigación, escuelas normales y centros de formación especializada). En muchas ocasiones a ello se añade otro obstáculo: la opacidad con que se trata la información sobre los enfoques evaluativos que desarrollan las universidades. De este modo, en palabras de Rueda y Luna «la finalidad de este estudio era que el conocimiento producido estuviera alejado de la información disponible en los documentos oficiales, regularmente centrada en los aspectos macrosociales; por lo general, expresada en indicadores numéricos y apegada a mostrar un cumplimiento de las políticas previamente establecidas» (Rueda y Luna 2001: 25-26).

RIED, con este estudio, intentó dinamizar la mejora acerca del desempeño docente, propiciando, con su modo de hacer, que el diagnóstico contribuyera a un conocimiento social compartido de las prácticas de evaluación del desempeño docente y con esto poner de manifiesto la necesidad de atender estas tareas para la mejora educativa en las universidades.

Los objetivos del estudio que aquí se presenta se centran en el modo en que se valora la función docente que realiza el profesorado y pueden sintetizarse en: a) identificar las características más relevantes de los procesos de evaluación de la docencia en las universidades objeto de estudio, b) elaborar un balance de las políticas relacionadas y c) establecer líneas de actuación para mejorar la evaluación del profesorado.

Participaron de manera voluntaria setenta y ocho universidades públicas mexicanas, seleccionadas teniendo en cuenta que se pudiera disponer de representatividad a nivel nacional en función de las regiones que establece ANUIES, basada en agrupaciones de estados. Asimismo, se tuvo en cuenta que hubiera representación de diversas tipologías de universidades respecto a su tamaño (pequeñas, medianas y grandes). En cada una de ellas, el informante era el responsable institucional de la evaluación del profesorado.

El diagnóstico se realizó a partir de la información recogida mediante tres estrategias: 1) cuestionario general (dirigido a recabar información acerca de los inicios de los procesos de la evaluación docente, las características principales de su aplicación y las consecuencias asociadas a los resultados de la evaluación), 2) guía para la elaboración de un ensayo (donde se solicita información acerca de las prácticas cotidianas de la evaluación y los elementos involucrados en la realidad de los procesos) y 3) seminario-taller con los responsables de los procesos de evaluación en cada una de las universidades.

En esta obra se aportan los resultados de este estudio. Se estructura en torno a los análisis realizados sobre las universidades públicas de cada una de las regiones y se aporta un marco de conclusiones en el

que se enfatizan las recomendaciones de mejora para estos procesos. Incluye, pues, ocho capítulos:

La valoración del desempeño docente en las universidades, donde se presentan los fundamentos y características del estudio, realizado por Mario Rueda y Edna Luna.

Los estudios de las instituciones por regiones, que se integran en seis capítulos: *Región metropolitana* (presentado por Mario Rueda, M^a Eugenia Linares y Paola M. Reyes), *Región Centro-Occidente* (Alma Figueroa), *Región Centro-Sur* (Leticia Elizalde, Brenda I. Olvera y Alma D. Torquemada), *Región Noroeste* (Edna Luna y Vicente Arámburo), *Región Noreste* (Benilde García-Cabrero y Aurora Bustillo) y *Región Sur-Sureste* (Javier Loredó y Edith J. Cisneros-Cohermour).

Análisis global de los resultados obtenidos y recomendaciones de mejora de los procesos de evaluación de la docencia universitaria para México; capítulo realizado por Mario Rueda, Edna Luna, Benilde García-Cabrero y Javier Loredó.

En *Anexos* se aportan instrumentos y guías utilizadas para la recogida de información.

En conjunto, se trata de un estudio de gran alcance, desarrollado con rigor, que intenta representar de la manera más fiel posible la realidad de los procesos evaluativos en las universidades de México. Por ello, entendemos que puede ser de gran utilidad a investigadores, evaluadores y docentes como referencia para el conocimiento de la situación mexicana y, por extensión, para la mejora del conocimiento acerca de los métodos y usos habituales de la evaluación de la docencia universitaria.

Carmen Carmona Rodríguez
Universitat de València

LEYVA, Y. E. (2012). *Evaluación de competencias médicas. Validez de constructo de pruebas criterios*. Alemania: LAP LAMBERT Academic Publishing GmbH & Co. KG, EAE: Editorial Académica Española, 295 páginas.

Yolanda E. Leyva Barajas es licenciada en Psicología con Maestría en Análisis Experimental de la Conducta por la UNAM y doctora en Educación por la Universidad de Aguascalientes (UAA) de México. Tiene una amplia trayectoria en medición y evaluación educativas y ha desarrollado su labor en instituciones como el Centro Nacional de Evaluación para la Educación Superior (CENEVAL) de México. Actualmente es directora de Evaluación del Instituto de Investigación en Tecnología Educativa (INITE) de México. Actúa como consultora de procesos de evaluación para diversos organismos de Educación Superior.

En este libro se discute el concepto de validez del constructo de una prueba nacional —el examen general para el egreso de la licenciatura en Medicina general (EGEL-MG) del Centro Nacional de Evaluación para la Educación Superior (CENEVAL)—, construido a partir de casos clínicos para evaluar competencias médicas de egresados de las carreras de Medicina de las principales universidades de México. La investigación desarrollada reúne aportaciones de la investigación en ciencia cognoscitiva a la solución de algunos problemas implicados en el desarrollo y validación de constructos. Tanto la teoría como la investigación en este campo ayudan en la comprensión del papel que juegan estructuras mentales tales como estrategias, procesos y representaciones así como sus respectivas interacciones en la ejecución de tareas cognoscitivas. De manera especial en la construcción y validación de las pruebas referidas a criterio, los investigadores se han orientado a tender un vínculo entre este campo de la investigación educativa y el campo de la investigación de la psicología cognoscitiva. La validación de pruebas referidas a criterio, bajo

esta perspectiva, requiere de la definición de diversas líneas de investigación en función del tipo de prueba, su estructura y los usos para los cuales se destina.

Siguiendo este enfoque, la autora exploró el constructo de la prueba que se eligió para establecer el proceso de validación, lo cual requirió ampliar el alcance de la investigación y el ámbito de estudio, el cual se expone en esta publicación. En el marco teórico se presenta una síntesis de las aportaciones teóricas derivadas de la investigación básica de la Psicología cognoscitiva al campo de la investigación aplicada en la evaluación de competencias médicas, incluyendo elementos relacionados con aportaciones al campo de la validez de constructo de pruebas educativas de competencias. Esta revisión del estado de conocimiento en el campo de los modelos mentales que subyacen a las competencias médicas y a su evaluación brindó la posibilidad de comprender la estructura del constructo y de identificar los procesos y estructuras mentales generalmente asociados a la solución de casos clínicos.

El libro está organizado en cuatro capítulos: en el primer capítulo se presentan el planteamiento del problema de investigación, los objetivos y las líneas de investigación para obtener evidencia de validez de las interpretaciones que se hacen a partir de las mediciones de la prueba, de acuerdo con los propósitos para los cuales fue construida.

El segundo capítulo contiene la exposición del marco teórico en cuatro apartados:

En el primero se presenta una base conceptual de la evaluación de competencias, los principales hallazgos de la investigación sobre el

dominio del conocimiento de ciencias biomédicas y de competencias clínicas que han contribuido al desarrollo de teorías cognoscitivas de aprendizaje y adquisición en la educación médica, así como una reseña de la evaluación de competencias en las profesiones de la salud a través de la solución de casos clínicos.

En el segundo apartado, se exponen los antecedentes y características de las pruebas referidas a criterio; describiendo los procedimientos de la definición y estructura del dominio por evaluar y una breve descripción de los principales métodos empleados para el establecimiento de estándares o puntos de corte para diferenciar niveles de competencia.

El tercer apartado presenta una breve reseña de la confiabilidad de una prueba criterial desde la teoría clásica y de la teoría de respuesta al ítem, así como una breve descripción de la teoría de la generalizabilidad.

El cuarto y último apartado aborda con mayor amplitud el problema de la validez y los diversos procedimientos de validación que se han empleado para pruebas referidas a criterio en la evaluación de competencias.

El tercer capítulo contiene información de los estándares internacionales de educación de Medicina: aquellos que se consideran requerimientos globales esenciales en la educación médica, así como una referencia a los principales estándares en uso. Se expone además el perfil referencial de validez, el contenido y la estructura de la prueba por validar, así como la metodología que sirvió para diferenciar los tres niveles de desempeño a partir de los resultados obtenidos mediante la prueba.

El cuarto capítulo presenta la metodología, organizando los estudios y análisis realizados en las tres líneas de evidencia de validez en torno a la prueba: análisis lógicos a partir de juicios de expertos, técnicas de análisis de datos empíricos y estudios con criterios externos. Para cada tipo de estudio se describen los procedimientos de recopilación y análisis de evidencias, los resultados obtenidos y una breve discusión del significado de los datos. En un primer intento por integrar la evidencia derivada de los estudios realizados se presenta un balance de aquellos aspectos de la prueba que contribuyen a la validez del constructo, así como aquellos que deben mejorarse para minimizar fuentes potenciales de invalidez.

El texto finaliza con las conclusiones, las cuales se integraron a partir de un análisis de la evidencia obtenida en torno a la naturaleza del constructo de la prueba a la luz de las teorías desarrolladas. Incorpora una reflexión del alcance del trabajo y las contribuciones metodológicas al estado de conocimiento en la construcción y validación de pruebas criteriosales a gran escala y concluye con un apartado en donde se presenta una propuesta de estudios futuros para profundizar en las diversas líneas de investigación generadas.

En síntesis, se trata de una aportación de gran interés e impacto, tanto para el diseño y desarrollo de instrumentos de evaluación de competencias de egreso como para el mejor conocimiento de las competencias clínicas.

Purificación Sánchez-Delgado
Universitat de València

LUNA SERRANO, E. (2011). *Aportaciones de la investigación a la evaluación de estudiantes y docentes*. Ciudad de México: Miguel Ángel Porrúa, 207 páginas.

Esta obra, coordinada por la doctora Edna Luna Serrano, constituye una excelente muestra de cómo la investigación aporta un cuerpo de conocimientos valioso para la evaluación educativa. El libro está dividido en dos grandes apartados. El primero de ellos se conforma por capítulos dedicados a la evaluación del aprendizaje en sí mismo, así como al estudio de las variables asociadas a este. El segundo apartado se centra en la evaluación de la docencia.

Por la calidad y cuidado con el que cada capítulo está escrito, así como por la aportación que cada uno hace a la obra, resulta ocioso decidir cuál de ellos se considera más importante o merecedor de mayor atención, pues todos en lo individual y en su conjunto son inestimables. Sin embargo, a partir del contenido de cada capítulo se puede identificar una estructura tácita que esquematiza la temática del libro.

En el apartado *Evaluación del aprendizaje*, se logra identificar tres rubros principales. En el primer rubro se agrupan capítulos que describen métodos y modelos para la evaluación del aprendizaje. Así por ejemplo, en el primer capítulo: *Método conceptual-empírico para determinar estándares de ejecución de la expresión escrita*, escrito por Eduardo Backhoff Escudero, Margarita Peón Zapata y Jesús Miguel Jornet Meliá, se describe el método para establecer niveles de logro y puntos de corte en la evaluación de la competencia de escritura. En el mismo rubro se puede ubicar el capítulo escrito por Luis Ángel Contreras Niño titulado *Modelos para desarrollar pruebas de gran escala de referencia criterial alineadas con el currículum*, donde se detalla una propuesta acerca de la sistematización del proceso para el desarrollo de pruebas criterios

alineadas a currículo y aplicadas a gran escala. Finalmente, en el capítulo *Examen de habilidades y conocimientos básicos (EXHCOBA): fortalezas y prospectiva*, Norma Larrazolo, Felipe Tirado y Martín Rosas describen las propiedades de este examen abordando su concepción, los procedimientos que utiliza y las características de su operación.

Se ubica de forma virtual un segundo rubro en el que se sitúan dos capítulos que describen estudios en los que, bajo distintas aproximaciones metodólicas, se analizan las variables asociadas al aprendizaje. Estos capítulos son: *Variables personales y escolares que afectan el rendimiento académico en la educación secundaria*, escrito por Joaquín Caso Niebla y Coral González Barbera y *Las oportunidades para aprender español y matemáticas: análisis multinivel aplicado a una muestra de estudiantes de tercero de secundaria en Baja California*, bajo la autoría de Juan Carlos Rodríguez Macías.

Un tercer rubro identificado se enfoca en los retos de la evaluación del aprendizaje. Este rubro es cubierto por el capítulo *¿Cómo se evalúa la calidad educativa de un país? Algunos desarrollos recientes*, en donde Eduardo Backhoff Escudero, Arturo Bouzas Riaño y Manuel González Montesinos, además de presentar un panorama general de la evaluación del logro educativo, muestran una serie de reflexiones acerca de las fortalezas y debilidades de la evaluación a gran escala, particularizando el caso de México.

Como preámbulo a la descripción del segundo apartado nombrado *Evaluación de la docencia*, es necesario mencionar que este representa uno de los varios atractivos del libro, pues es común que, cuando

se lee *evaluación educativa*, el lector evoca casi de forma inmediata la evaluación del aprendizaje en los estudiantes, descontando la relevancia de la evaluación docente. Sin embargo, en la obra se rescata este punto a través de dos capítulos, por demás interesantes. Graciela Cordero Arrollo, Edna Luna Serrano, Rosa María Pons Parra y José Manuel Serrano González-Tejero en el capítulo *La evaluación de los programas de formación de profesores: el caso de un programa en métodos de aprendizaje cooperativo en una universidad pública estatal*, enuncian los retos que se enfrentan en la evaluación formativa y sumativa de los programas de formación de profesores universitarios. En el mismo apartado, el capítulo *La efectividad de la enseñanza y su relación con las características del profesor y del*

curso: aportaciones para el diseño de sistemas de evaluación docente, escrito por Edna Luna Serrano, Vicente Arámburo Vizcarra y Graciela Cordero Arroyo se analiza cómo las características del docente afectan a la evaluación de su desempeño y cómo estas a su vez se asocian con las características del curso.

En definitiva, cada capítulo ofrece información derivada de estudios y análisis minuciosos relacionados con diferentes aspectos de la evaluación de estudiantes y docentes. Por lo tanto, además de útil e interesante, resulta una lectura obligada para todos aquellos interesados en el tema de la evaluación educativa.

Alicia A. Chaparro Caso López
 Universidad Autónoma de Baja
 California (México)

BOUZAS, A., ELBITTAR, A., PALAFOX, G., PEÑA, P., LEENEN, I., MONROY, L., HERRERA, M., GARCÍA, I. (2012). *Plan de desarrollo del sistema integral de cuestionarios de contexto del Ceneval*. México: Ceneval, A. C.

El Centro Nacional de Evaluación para la Educación Superior, A.C. (Ceneval) surgió en México en 1994 —por mandato de la Asociación Nacional de Universidades e Instituciones de Educación Superior— ante la necesidad de contar con una instancia de evaluación externa que promueva la mejora de la calidad educativa mediante la generación de información fiable y útil para tomar decisiones de alto impacto educativo.

En aras de cumplir con su misión, el Ceneval no se limita a construir exámenes que miden los aprendizajes logrados por los individuos, sino que también desarrolla un conjunto de instrumentos complementarios —los cuestionarios de contexto— que exploran una amplia gama de variables personales, escolares, económicas y sociales de los sustentantes de las pruebas de logro académico.

Los cuestionarios de contexto permiten conocer las características de la población sustentante, comparar generaciones y contextualizar las diferencias observadas en los puntajes de los exámenes. Adicionalmente, los datos conjuntos del logro académico y los del contexto de los sustentantes posibilitan el desarrollo de múltiples líneas de investigación que, por supuesto, incluyen modelos explicativos de los factores que impactan en el aprendizaje

En fechas recientes, el Ceneval se dedicó a la tarea de consolidar y normar los lineamientos institucionales para construcción de los cuestionarios de contexto. Ante este reto, se convocó a un grupo de investigadores de reconocido prestigio que, con sus aportaciones colegiadas y continua intervención, permitieron elaborar el *Plan de desarrollo del sistema integral de cuestionarios de contexto*.

En esta publicación se explica que el sistema integral de cuestionarios de contexto del Ceneval está conformado por instrumentos de medición que se aplican a los sustentantes de exámenes de ingreso, de egreso de licenciaturas, de diagnóstico formativo y sumario, de acreditación del bachillerato y a sustentantes de exámenes de certificación de diferentes profesiones.

En el *Plan de desarrollo del sistema integral de cuestionarios de contexto*, el Ceneval hace explícito el propósito de cada cuestionario, detalla el diseño modular que subyace a cada tipo de instrumento, explica la metodología que se implementa para su construcción, presenta los métodos estadísticos que se utilizan para analizar las escalas y detalla

el marco de referencia que subyace a las variables del módulo básico que une transversalmente a los cuestionarios.

El *Plan de desarrollo del sistema integral de cuestionarios de contexto* es un libro útil para guiar a investigadores en el diseño y construcción de escalas específicas, cuestionarios o encuestas.

Es importante señalar que el libro se encuentra en la fase final de la revisión editorial, por lo que su publicación se tiene contemplada para el mes de junio. Por lo pronto, la versión electrónica puede ser consultada en el portal electrónico del Ceneval.

Javier Díaz de la Serna
Ceneval, A. C. México