

Evidencias de validez del Cuestionario de Evaluación de la Competencia Docente Universitaria

ERIKA PAOLA REYES PIÑUELAS* | EDNA LUNA SERRANO**
JOAQUÍN CASO NIEBLA***

El propósito del artículo es presentar la estrategia metodológica seguida para la obtención de evidencias de validez de la estructura interna del Cuestionario de Evaluación de la Competencia Docente (CECD) con base en la opinión de los estudiantes. El CECD se aplicó a muestras representativas de estudiantes de licenciatura de dos universidades mexicanas ($n_1=511$ y $n_2=503$) y se utilizó el análisis factorial confirmatorio como procedimiento analítico. La modelización confirmatoria incluyó la identificación del modelo, la especificación del modelo, la selección de medidas, la evaluación de los criterios de calidad de ajuste y la modificación del modelo de medición. Los resultados permitieron confirmar el ajuste del modelo hipotético con el sustento del CECD; se observó una estructura bifactorial con índices de ajuste aceptables e índices de consistencia interna superiores a .90 en ambas muestras. El sustento teórico y empírico del CECD avala su utilización en los ámbitos universitarios descritos.

The purpose of this article is to present the methodological strategy used to obtain evidence of validity of the internal structure of the University Teaching Skills Evaluation Questionnaire (Spanish acronym CECD) based on students' opinions. The CECD was applied to representative samples of undergraduate students at two Mexican universities ($n_1=511$ and $n_2=503$) and confirmatory factor analysis was used as the analytical procedure. The confirmatory modeling included identification of the model, specification of the model, selection of measurements, evaluation of criteria of goodness of fit, and modification of the measurement model. The results helped to confirm the fit of the hypothetical model with the support of the CECD, and we observed a bifactor structure with acceptable indices of fit and indices of internal consistency above .90 in both samples. The theoretical and empirical foundation of the CECD supports its use in the university settings described.

Palabras clave

Evaluación
Competencias
Validez de las pruebas
Docencia
Práctica docente
Instrumentos de medición

Keywords

Evaluation
Competencies
Validity of tests
Teaching
Teaching practice
Measuring instruments

Recepción: 25 de junio de 2018 | Aceptación: 13 de diciembre de 2019

DOI: <https://doi.org/10.22201/iisue.24486167e.2020.169.58931>

- * Profesora-investigadora de la Facultad de Ciencias Humanas de la Universidad Autónoma de Baja California (UABC) (México). Doctora en Ciencias Educativas. Línea de investigación: evaluación educativa. Publicación reciente: (2020, en coautoría con E. González), "Adaptación de un cuestionario de habilidades para la vida en adolescentes", en E. Vargas, S. Cruz y V. López (coords.), *Interdisciplina en temas de relevancia social*, México, Porrúa, pp. 37-59. CE: erikareyes@uabc.edu.mx
- ** Investigadora del Instituto de Investigación y Desarrollo Educativo (IIDE) de la Universidad Autónoma de Baja California (UABC) (México). Doctora en Educación. Línea de investigación: evaluación educativa. Publicación reciente: (2020, en coautoría con A. Hernández), "Desarrollo de un cuestionario de evaluación de la competencia docente en línea", *Revista Iberoamericana de Educación a Distancia*, vol. 23, núm. 2, pp. 307-328. DOI: <https://doi.org/10.5944/ried.23.2.27084>. CE: eluna@uabc.edu.mx
- *** Investigador del Instituto de Investigación y Desarrollo Educativo (IIDE) de la Universidad Autónoma de Baja California (UABC) (México). Doctor en Psicología Educativa. Líneas de investigación: evaluación del aprendizaje; diseño de pruebas en contextos educativos; evaluación de convivencia escolar; evaluación de prácticas y procesos institucionales en educación superior. Publicación reciente: (2020, en coautoría con J. Moreno y C. Díaz), "Diagnóstico de competitividad académica y acreditación de programas educativos de la Universidad Autónoma de Baja California, México", *Archivos Analíticos de Políticas Educativas*, vol. 28, núm. 22. DOI: <https://doi.org/10.14507/epaa.28.4598>. CE: jcaso@uabc.edu.mx

INTRODUCCIÓN

La evaluación docente como práctica social se considera un recurso fundamental para la mejora de los procesos educativos; pero, al mismo tiempo, se reconoce que no posee características intrínsecas que aseguren su bondad. Un programa inadecuado de evaluación puede tener efectos nocivos en la comunidad educativa; por ello, se advierte una serie de elementos a considerar en el diseño de programas de evaluación, entre los que destacan la explicitación de los propósitos y repercusiones de la evaluación, y la consideración de la complejidad de la enseñanza y las particularidades del contexto institucional (RIIED, 2008; Van Der Schaaf *et al.*, 2019).

Entre las estrategias utilizadas para evaluar la docencia destacan los cuestionarios de evaluación docente con base en la opinión de los estudiantes, en virtud de que cuentan con más de 80 años de investigación que respalda su empleo casi universal en el ámbito universitario (Spooren *et al.*, 2013). Algunas de las razones a favor de su uso generalizado son: la posibilidad de una amplia cobertura de estudiantes facilitada por el uso de la tecnología en su aplicación, su bajo costo y el ágil procesamiento de la información.

Si bien los puntajes derivados de los cuestionarios se consideran válidos, confiables y relativamente libres de sesgo (Linse, 2017; Marsh, 2007; Van Der Schaaf *et al.*, 2019), persisten críticas respecto a su utilidad tanto en evaluaciones formativas como sumativas. Se cuestiona su validez, dado que la mayoría de los cuestionarios en uso no han incorporado los cambios contextuales particulares de las situaciones educativas, en específico las nuevas prácticas instruccionales derivadas de los planes de estudio por competencias, aprendizaje colaborativo y la educación a distancia (Spooren *et al.*, 2013). En este sentido, se reconoce que incorporar la aproximación de enseñanza por competencias en los programas de estudio conlleva una concepción de

aprendizaje que demanda de la práctica docente características específicas, las cuales deben reflejarse en el contenido de las dimensiones y reactivos de los cuestionarios. Además, al reconocer el gran poder orientador que tienen los procesos de evaluación docente persiste la necesidad de contar con instrumentos válidos, confiables y pertinentes a los distintos contextos educativos, con el fin de obtener información congruente que posibilite mejorar las experiencias de enseñar y aprender.

En un ejercicio por superar las limitaciones en el desarrollo y validación de los instrumentos de evaluación docente, el objetivo general de este trabajo es describir la experiencia de acumulación de evidencias de validez basadas en la estructura interna del Cuestionario de Evaluación de la Competencia Docente (CECD), alineado al modelo de enseñanza por competencias y a los nuevos estándares publicados en el 2014 por la American Educational Research Association (AERA), la American Psychological Association (APA) y el National Council on Measurement in Education (NCME).

Los objetivos específicos son:

- Confirmar que el modelo hipotético-deductivo del CECD se ajusta a los datos derivadores de su aplicación a muestras de estudiantes de la UADY y UACH.
- Comprobar que las dimensiones del Modelo de Evaluación de la Competencia Docente (MECD) se encuentren representadas en la estructura interna del CECD.

CUESTIONARIOS DE EVALUACIÓN DOCENTE CON BASE EN LA OPINIÓN DE LOS ESTUDIANTES

De manera frecuente, los puntajes de los cuestionarios son utilizados por los docentes como fuente de información sobre su enseñanza; por los estudiantes, como un criterio a considerar en la selección de cursos, y por los administradores como un indicador para

la toma de decisiones administrativas (Hornstein *et al.*, 2017). En particular respecto al uso que los alumnos hacen de cuestionarios de evaluación de la docencia, las buenas prácticas de evaluación sugieren la necesidad de explicitar los usos de los resultados y velar por la credibilidad del proceso de evaluación (Benton y Li, 2015; Royal, 2017).

De acuerdo con House (1998), la credibilidad de las evaluaciones se fundamenta en dos elementos prácticos: 1) la utilización de una metodología científica objetiva; y 2) la independencia política de los evaluadores respecto de los programas y políticas evaluadas. En este sentido, en el caso particular de la evaluación de la docencia, es indispensable contar con instrumentos de calidad métrica que midan los elementos centrales del constructo competencia docente, así como con procesos de evaluación dirigidos por personal capacitado independiente de la comunidad educativa.

Cabe señalar que la teoría sobre la enseñanza que sustenta a la mayoría de los cuestionarios de evaluación docente por los alumnos considera que la enseñanza efectiva es multidimensional, por lo que se propone evaluar las diferentes dimensiones que la conforman. Por lo general, la agrupación de los reactivos es producto de los análisis factoriales exploratorios y/o confirmatorios a los que fueron sometidos los instrumentos, y refleja las dimensiones de la enseñanza a juicio de los estudiantes (Boysen, 2016). Además, a partir de los estudios de correlación entre los puntajes de los cuestionarios y el aprovechamiento de los estudiantes las dimensiones asociadas a la efectividad de la enseñanza se han clasificado, en orden de importancia, en: claridad y entendimiento, estimulación del interés por la materia, preparación y organización del curso, logro de los objetivos planteados y motivación a los estudiantes para tener alto rendimiento (Feldman, 2007).

Simultáneamente, se asume que las dimensiones de la enseñanza efectiva y los reactivos que las incluyen forman parte de

las habilidades genéricas que todo docente debe presentar. En el marco de esta lógica se encuentran los cuestionarios típicos y ampliamente reconocidos, como es el caso del Student Evaluation of Educational Quality (SEEQ) diseñado por Marsh en 1982, que se ha utilizado tanto en el ámbito anglosajón como en el europeo y latinoamericano (Marsh, 2007; Andrade-Abarca *et al.*, 2018). No obstante, desde principios de los años noventa se habla de las limitaciones de incluir únicamente las dimensiones asociadas a la efectividad docente y la necesidad de incorporar los cambios contextuales propios de cada situación educativa, como una manera de considerar el contexto pedagógico y de responder a la filosofía y misión de la institución (Luna, 2010).

En este sentido, la adopción de la modalidad de enseñanza por competencias involucra un cambio en las prácticas educativas y, por consecuencia, en su evaluación; en principio, por lo tanto, resulta obligado que el contenido de los instrumentos de evaluación sea coherente con el modelo de enseñanza (Cano, 2015). Por otra parte, se reconoce que el desarrollo de los instrumentos de evaluación de la docencia por los estudiantes debe incluir tres estrategias: a) una revisión de los componentes principales de la efectividad docente aunado a la retroalimentación de profesores y estudiantes; b) la aplicación de métodos empíricos como el análisis factorial y el análisis multirrasgo-multimétodo; y c) su fundamentación en una teoría de la enseñanza (Marsh, 2007). Las dos últimas estrategias coinciden también con las pautas sugeridas por Van Der Schaaf *et al.* (2019); en estos casos los componentes teóricos de la enseñanza deben provenir del modelo de enseñanza por competencias.

En una revisión de 127 instrumentos reportados en publicaciones entre 2000 y 2016, Van Der Schaaf *et al.* (2019) señalan, como práctica común en el desarrollo de los instrumentos de medición de la práctica docente, que las evidencias de medidas de confiabilidad y validez en la rúbrica de evaluación de la calidad

de los instrumentos utilizada resultaron con una calidad *media*, tanto en las evidencias de confiabilidad como de validez, en el 66 y 76 por ciento de los instrumentos analizados respectivamente, de lo que se concluye la necesidad de evidencia teórica y empírica de calidad. En el contexto mexicano, Moreno (2018) señala la ausencia de un fundamento teórico sólido como una limitante en los trabajos de evaluación de la docencia en educación superior.

Lo anterior se enmarca en el tema de la validez, que no es un atributo de los instrumentos de medición en sí mismos, sino que se encuentra en las interpretaciones de sus puntuaciones; se trata de un proceso de acumulación de pruebas sobre el grado en que las evidencias empíricas y la teoría apoyan las interpretaciones, y se refiere a la configuración de un argumento de validez con solidez científica que respalda la interpretación de las puntuaciones en el instrumento de medición y su importancia para el uso que fue desarrollado. Así mismo, la validez tendrá significado en tanto se circunscriba a qué interpretación o tipo se refiere, ya que ambos aspectos forman un continuo; en términos generales se puede decir que es una cuestión de grado, más que una propiedad que implique ausencia o presencia (Messick, 1995). En este sentido, una prueba muestra un grado aceptable de validez para ciertos usos específicos y con ciertas poblaciones.

Por otro lado, la presencia insuficiente de teoría que sustente al instrumento no sólo se presenta en la evaluación de la docencia. Borsboom (2006) refiere que en el ámbito de la psicometría la fundamentación teórica de los instrumentos es escasa, dado que una práctica generalizada en el desarrollo de los instrumentos es prestar mayor atención a los análisis de la prueba y posteriormente averiguar qué se está midiendo, cuando el proceso debiera ser a la inversa; o bien, se utilizan cuestionarios de opinión de manera sistemática sin contar con evidencias de validez o sin probarse en el contexto que son utilizados (Spooren *et al.*, 2013).

Asimismo, en el tratamiento estadístico de los datos se observa que éstos son del tipo de evaluaciones que hemos descrito, las cuales suelen utilizar metodologías poco robustas que subestiman el papel del sustento teórico como punto de partida de todo proceso de validación empírica (Van Der Schaaf *et al.*, 2019). Para la acumulación de evidencias basadas en la estructura interna, Rios y Wells (2014) señalan que, en este tipo de evidencias de validez, algunos de los aspectos que se abordan son la dimensionalidad y la confiabilidad. El propósito que se persigue al evaluar la dimensionalidad es determinar si las interrelaciones entre los ítems apoyan las puntuaciones de los exámenes que se van a utilizar para extraer inferencias, mientras que, en el caso de los índices de fiabilidad, éstos proporcionan evidencia de que las calificaciones de los resultados reportados son consistentes a través de la repetición en la administración de las pruebas (Benton *et al.*, 2015). El análisis de factores es un método estadístico común utilizado para evaluar la dimensionalidad de un conjunto de datos. A este respecto, el modelamiento de ecuaciones estructurales (MEE) representa un procedimiento analítico idóneo, pues permite contrastar los supuestos teóricos y las formulaciones hipotéticas que fundamentan el diseño de este tipo de instrumentos.

La revisión de la literatura permitió identificar tres estudios que documentan el diseño, desarrollo y uso de cuestionarios basados en un enfoque por competencias para evaluar la práctica docente, uno con base en la opinión de los estudiantes y dos de autoevaluación. Córdova *et al.* (2015) reportan el diseño de un cuestionario orientado a la evaluación de las competencias que deben proveerse mediante la enseñanza de asignaturas preclínicas en la licenciatura de Medicina. Dicho instrumento consta de seis apartados: sobre la universidad, la asignatura, la relación con el profesor, la metodología y las competencias generales. Registra índices de consistencia interna de .92 y 60 por ciento de varianza explicada.

En lo relativo a los cuestionarios de autoevaluación, Gargallo *et al.* (2011) plantean la evaluación de la metodología docente y evaluativa de los profesores universitarios a partir de tres competencias: una a nivel de planificación, otra de actuación/interacción en clase y fuera de clase con sus alumnos, y otra de evaluación. Los autores documentaron índices aceptables de consistencia interna (.83) del instrumento y evidencias de validez de su estructura interna. Por su parte, Bujanda (2017) documenta el desarrollo de un cuestionario de competencias docentes del área de Servicios Socioculturales y Comunidad que involucra las competencias de docentes de las licenciaturas en Educación social, Trabajo social, Psicología, Pedagogía y Psicopedagogía. El instrumento se conforma por nueve dimensiones y cuenta con evidencia de validez de contenido, así como índices de consistencia interna altos (.96).

Por otro lado, un cuestionario de referencia obligada por los estudios que le preceden y la solidez de sus medidas es el SEEQ, el cual, de acuerdo con un estudio de Andrade-Abarca *et al.* (2018), registra un índice de confiabilidad de .97 y un 79.6 por ciento de varianza explicada.

Modelamiento de ecuaciones estructurales (MEE)

Por su carácter de tipo confirmatorio, esta metodología permite utilizar datos empíricos para evaluar la validez del modelo teórico que es explicitado a priori a partir de la teoría que lo fundamenta. El MEE resulta robusto para la obtención de evidencias de validez de constructo donde se ubican las evidencias basadas en la estructura interna; de acuerdo con los estándares para el desarrollo de pruebas, hace referencia a “analizar el grado en que las relaciones entre los reactivos de una prueba y sus componentes corresponden a la estructura de dimensiones del constructo medido” (AERA-APA-NCME, 2014: 16).

El MEE pertenece a la familia de estadísticos multivariantes y permite la contrastación de

modelos hipotéticos que proponen asociaciones de causa-efecto y relaciones entre variables (Rios y Wells, 2014). Su robustez y amplia variedad de programas de procesamiento de datos ha hecho de esta metodología una alternativa accesible para el desarrollo de modelos complejos y teorías explicativas que conciernen al comportamiento humano (Ruiz *et al.*, 2010).

Tres situaciones en las que el MEE se diferencia de otras técnicas son: reconoce errores de medición e identifica si una relación entre variables es significativa o es débil; revela la existencia de una interdependencia o causalidad simultánea entre las variables observadas; e identifica cuando no se han observado importantes variables explicativas (Hair *et al.*, 2010).

Los pasos sugeridos para la aplicación de esta técnica, según Kline (2016), son: especificación del modelo; identificación del modelo; selección de medidas, recopilación, preparación y selección de los datos; evaluación del modelo; re-especificación del modelo; informe de resultados. A continuación se detallan estos pasos:

1. *Especificación del modelo.* Se expresa la estructura factorial, con las combinaciones de las relaciones entre variables latentes y variables observadas.

2. *Identificación del modelo.* Los modelos estadísticos generalmente deben respetar ciertas reglas o restricciones; un requisito es el de la identificación. Para identificar un modelo, Guardia (2016) señala que deben calcularse los grados de libertad existentes y asociados al modelo que se propone. El modelo es identificado si los grados de libertad resultan cero o mayor a cero y se procede al siguiente paso; si es menor a cero el modelo no está identificado y su estimación no es posible.

3. *Selección de medidas, recopilación, preparación y selección de los datos.* Indicar la confiabilidad de la muestra e informar sobre los coeficientes de otras muestras y detallar si son similares a las que se reportan; informar sobre

la unidimensionalidad de cada uno de los factores planteados en el modelo a evaluar. Además, informar sobre los datos como pérdida de datos sistemática o aleatoria, y la elección de algún método de imputación de datos perdidos y tipo de diseño de muestreo probabilístico. Informar, así mismo, sobre estadísticos descriptivos, valores de índices de sesgo y curtosis. Finalmente, indicar el tipo de matriz de datos analizada.

4. *Evaluación del modelo.* Para este paso: a) se evalúa el ajuste del modelo, es decir, se determina qué tan bien explica los datos y, si no lo hace bien, se valora si pasar al siguiente paso (reespecificación del modelo) es justificable dada la teoría que subyace y los resultados de estudios empíricos previos; b) si se conserva un modelo, interpretar las estimaciones de los parámetros (omitir el paso 6); c) considerar un modelo equivalente o casi equivalente.

La estimación del modelo puede realizarse mediante dos métodos: el primero es el ajuste exacto, el cual responde a una pregunta dicotómica y se evalúa usando el test chi cuadrada con grados de libertad donde los valores de significación estadística mayores que 0.05 o 0.01 indican que el modelo propuesto se ajusta bien a las varianzas y covarianzas observadas.

Por tradición se ha evaluado el ajuste del modelo a partir de la medida de chi cuadrada, sin embargo, se ha documentado, entre otros problemas (Byrne y Van de Vijver, 2010; Hair *et al.*, 2010; Hu y Bentler, 1999), la sensibilidad de éste al tamaño de muestra, por lo que es posible que indique diferencias significativas en modelos que son equivalentes. Aunado a ello, normalmente se utiliza un conjunto de estadísticos de ajuste, puesto que ninguno de ellos aporta la información necesaria para valorar el modelo en su totalidad (Schreiber *et al.*, 2006).

El segundo método para evaluar el modelo es el ajuste aproximado, que consiste en calcular el grado en el que la hipótesis nula es aceptada o rechazada dentro de un continuo con diversos índices de ajuste (Kim y Bentler, 2006). Algunos de estos índices superan la sensibilidad de la medida de chi cuadrada; es el caso del Índice de Bondad de Ajuste Comparativo (CFI por sus siglas en inglés). Byrne y Van de Vijver (2010) sostienen que es un índice que presenta una menor sensibilidad al tamaño de la muestra y provee medidas de mayor confianza. En la Tabla 1 se presentan algunos de estos índices con la expectativa que establecen los autores considerados. Los índices se clasifican en índices de ajuste absoluto, índices de ajuste incremental e índices de ajuste de parsimonia.

Tabla 1. Criterios de corte de diversos índices de bondad de ajuste

Estadístico	Criterios de corte				
	Abrev.	Arbuckle 2013	Hair <i>et al.</i> 2010	Hooper <i>et al.</i> 2008	Ruíz <i>et al.</i> 2010
Ajuste absoluto		—		Valor pequeño χ^2	
Chi cuadrada	χ^2		$p=.05$	$p=.05$	$p=>.05$
Grados de libertad	gl				
Significación	p				
Chi cuadrada normado	χ^2/gl	—	<3.0	2.0-5.0	< 3.0
Raíz del residuo cuadrático promedio de aproximación	RMSEA	$\leq .08$.03-.08	<.08	< .08
Índice de bondad de ajuste	GFI	≤ 1.0	.95	$\geq .95$	$\geq .95$
Índice de bondad de ajuste corregido	AGFI	1.0	—	$\geq .95$	$\geq .95$
Raíz del residuo cuadrático promedio	RMR	0	—	Valor pequeño	Cercano a 0

Tabla 1. Criterios de corte de diversos índices de bondad de ajuste

(continuación)

Estadístico	Criterios de corte				
	Abrev.	Arbuckle 2013	Hair <i>et al.</i> 2010	Hooper <i>et al.</i> 2008	Ruiz <i>et al.</i> 2010
Raíz del residuo cuadrático promedio estandarizado	SRMR	—	<.1	<.05	—
Ajuste incremental Índice de Tucker-Lewis	TLI (NNFI)	>.90	Cercano a 1.00	>.95	≥ .95
Índice de ajuste normalizado	NFI	>.90	Cercano a 1.00	>.95	≥ .95
Índice de bondad de ajuste comparativo	CFI	1.00	1.00	>.95	≥ .95
Ajuste parsimonioso Índice de ajuste normado de parsimonia	PNFI	—	—	>.50	Cercano a 1.00

Nota: las rayas son índices de ajuste no reportados.

Fuente: elaboración propia.

Hu y Bentler (1999) reportaron algunos problemas en la evaluación global del modelo con chi cuadrada y plantearon una valoración de modelos de medición con una estrategia que consiste en combinar criterios de corte de dos índices de bondad de ajuste, como alternativa a los criterios convencionales. El ajuste del modelo se valora con las reglas propuestas,

de tal forma que se requiere de un valor de raíz del residuo cuadrático promedio de aproximación (RMSEA por sus siglas en inglés) de .06 o menor, y un valor de raíz del residuo cuadrático promedio estandarizado (SRMR por sus siglas en inglés) de .09 o menor para considerar un modelo con buen ajuste. En la Tabla 2 se muestra esta propuesta.

Tabla 2. Estrategia de presentación de dos índices de ajuste según Hu y Bentler (1999)

Combinación de índices	Regla
NNFI (TLI) y SRMR	NNFI de 0.96 o mayor y un SRMR de .09 o menor
RMSEA y SRMR	RMSEA de 0.06 o menor y un SRMR de 0.09 o menor
CFI y SRMR	CFI de .96 o mayor y un SRMR de 0.09 o menor

Fuente: elaboración propia.

Posteriormente, Kim y Bentler (2006) señalan que para evaluar el ajuste de los datos se recomienda el uso de por lo menos dos tipos de índices diferentes, ya sea de ajuste absoluto, incremental o parsimonia, por lo que se considera que un modelo presenta un buen ajuste cuando están presentes valores que cumplen la expectativa en más de una de estas categorías.

5. *Reespecificación del modelo.* Es frecuente que los modelos no ajusten de manera óptima de inicio debido a la complejidad de sus relaciones, además de las restricciones implícitas y explícitas que incluyen. Cuando sucede esto se propone la aplicación de recursos de revisión que permiten reespecificar las relaciones postuladas en el modelo original, y con ello, la mejora del ajuste con el modelo teórico.

Batista y Coenders (2012) plantean tres recomendaciones para realizar las modificaciones del modelo: a) examinar los resultados cada vez que se introduzca una modificación; b) introducir modificaciones que mejoran los estadísticos de ajuste al añadir parámetros y/o suprimir variables poco válidas para un modelo parsimonioso; c) realizar las modificaciones del modelo con base en una estrategia conjunta entre la interpretación teórica y el aporte de los valores estadísticos a la mejora del modelo. El último punto es de especial relevancia dado que las decisiones deben tomar en cuenta la teoría que subyace al modelo y su representatividad.

Muñiz y Fonseca-Pedrero (2008) señalan no perder de vista que el propósito de los análisis estadísticos debe ser siempre maximizar o potenciar las propiedades métricas del instrumento de medida, dado que no existen reglas o rutas universales y las consideraciones estadísticas no garantizan unos resultados con significado conceptual, por lo que deberían tenerse presentes también los aspectos sustantivos. En el mismo sentido, Hair *et al.* (2010) y Kim y Bentler (2006) señalan cómo saber, en el caso del ajuste de los datos con base en el estadígrafo χ^2 , si es sensible al tamaño de muestra; y Byrne y Van de Vijver (2010) añaden que esta situación representa, en ocasiones, una vía que pone en riesgo la representatividad de la teoría que fundamenta los instrumentos.

Finalmente, con respecto a la evaluación del ajuste del modelo cabe señalar que, aun

cuando el modelo propuesto presente un ajuste aceptable, éste representa uno de los modelos posibles que también pueden ajustarse a los datos (Kline, 2016; Ruíz *et al.*, 2010).

6. *Informe los resultados.* Al finalizar el análisis se deben de organizar los resultados estadísticos para que puedan ser informados.

CUESTIONARIO DE EVALUACIÓN DE LA COMPETENCIA DOCENTE CON BASE EN LA OPINIÓN DE LOS ESTUDIANTES

El CECD con base en la opinión de los alumnos fue desarrollado a partir de las competencias e indicadores del Modelo de Evaluación de la Competencia Docente (MECD) para la educación media y superior (García-Cabrero *et al.*, 2014). Este modelo contempla tres momentos o dimensiones de la práctica docente: 1) la previsión del proceso enseñanza aprendizaje incluye las creencias y conocimientos del profesor sobre la enseñanza, la planeación de su clase, las expectativas que posee del grupo y su eficacia como docente; 2) la conducción del proceso enseñanza aprendizaje considera las teorías en uso aplicadas por los profesores en la conducción didáctica, el conocimiento de los niveles de desempeño de los alumnos, la aplicación de estrategias de enseñanza y la generación de un clima social que favorezca el aprendizaje; y 3) la valoración del proceso enseñanza y aprendizaje involucra el uso de estrategias adecuadas para valorar el proceso

Tabla 3. Competencias del Modelo de Evaluación de la Competencia Docente en educación media y superior

Dimensiones	Competencia
Previsión del proceso enseñanza - aprendizaje	Planear el curso de la asignatura
Conducción del proceso enseñanza - aprendizaje	Gestionar la progresión de los aprendizajes (plan de clase)
	Llevar a cabo la interacción didáctica en el aula
	Utilizar formas de comunicación adecuadas para apoyar el trabajo académico
Valoración del impacto del proceso de enseñanza	Utilizar formas adecuadas para valorar los procesos de enseñanza aprendizaje, así como su impacto

Fuente: elaborado con información de García-Cabrero *et al.* (2014).

de enseñanza y aprendizaje. En la Tabla 3 se muestran las competencias que derivan de cada uno de los momentos mencionados.

Los principios que orientan el modelo son: una orientación de tipo formativa que favorece la reflexión y retroalimentación de la práctica docente, así como el planteamiento de acciones para el perfeccionamiento; orientación participativa, esto es, el docente y los directivos intervienen en el diseño e implementación; orientación humanista, según la cual el docente es considerado un individuo con intereses, motivaciones y emociones, y se busca cuidar su autoestima y dignidad; y un enfoque multirreferencial que privilegia la diversidad de opiniones y, además, parte de la complejidad de la evaluación de la docencia, con especial interés en los elementos que la determinan (García-Cabrero *et al.*, 2008a; García-Cabrero *et al.*, 2008b).

En el MECD se asume que las competencias pueden ser evaluadas de manera separada con base en la explicitación de criterios e indicadores (Denyer *et al.*, 2007). Esta perspectiva es

diferente a la de competencia global, integradora, que sólo puede evaluarse en contextos auténticos y globalizados.

Como parte del proceso de consolidación del MECD se han realizado diversos trabajos de investigación con el objetivo de contribuir al conocimiento de la docencia y su evaluación; estos estudios se realizaron en universidades de diferentes regiones de México, con enfoques metodológicos de tipo tanto cualitativo como cuantitativo, con la participación activa de docentes, directivos y estudiantes, entre otros, orientados a la definición de las competencias docentes, el desarrollo de instrumentos, y la promoción y reflexión a través de los portafolios, por citar algunos reportados principalmente en Cisneros-Cohernour *et al.* (2012) y en García-Cabrero *et al.* (2014).

El cuestionario desarrollado para la investigación que se describe aquí consta de 25 reactivos (Tabla 4) de respuesta tipo Likert con cinco niveles de respuesta: 1, nunca; 2, casi nunca; 3, algunas veces; 4, casi siempre; y 5, siempre.

Tabla 4. Cuestionario de Evaluación de la Competencia Docente con base en la opinión de los estudiantes

Ítems	1	2	3	4	5
Planeación y gestión del proceso de E-A					
1. Explica la utilidad de los conocimientos en situaciones de la vida diaria					
2. Puntualiza los conceptos importantes de los temas vistos en el curso					
3. Enseña procedimientos para usar el conocimiento en la solución de problemas reales					
4. En las primeras clases identifica los conocimientos de los estudiantes sobre la materia					
5. Modifica las actividades o fuentes de información para adecuarlas a las necesidades del grupo					
6. Plantea situaciones con problemas reales que favorecen mi aprendizaje de los temas					
7. Realiza actividades en clase que me motivan a aplicar lo aprendido fuera del aula					
8. Me motiva a continuar aprendiendo sobre los temas vistos en el curso					
9. Aborda los temas con una secuencia razonada					
Conducción y valoración del proceso de E-A					
10. Presenta la lista de temas y su secuencia en el programa de la materia					
11. Explica los criterios de evaluación de la materia					
12. El tiempo dedicado para aprender cada tema es suficiente					
13. Acepta sugerencias de los estudiantes sobre la manera de organizar las actividades					
14. Da oportunidades iguales a todos los estudiantes para participar en clase					

Tabla 4. Cuestionario de Evaluación de la Competencia Docente con base en la opinión de los estudiantes

(continuación)

Ítems	1	2	3	4	5
15. Estructura tareas de aprendizaje que propician la colaboración entre los estudiantes					
16. Se asegura de que en las actividades en equipo, cada integrante contribuya a su realización					
17. Cuando habla, expresa claramente sus ideas					
18. Sus exposiciones se caracterizan por tener un inicio, desarrollo y conclusión					
19. Fomenta la participación del grupo para realizar conclusiones en sus exposiciones					
20. Demuestra ser responsable en su trabajo					
21. Promueve la elaboración de acuerdos de convivencia en el aula con base en las opiniones de todo el grupo					
22. Es respetuoso con todos los estudiantes del grupo					
23. Realiza evaluaciones para retroalimentar sobre las dificultades de aprendizaje					
24. Respeto los criterios de evaluación planteados al inicio del curso					
25. La evaluación final fue congruente con los contenidos y actividades del curso					

Nota: 1=nunca; 2=casi nunca; 3=algunas veces; 4=casi siempre; y 5=siempre.

Fuente: elaboración propia.

El desarrollo y validación del CECD se reportó en un trabajo previo (Luna y Reyes, 2015). El estudio documenta la aplicación de este instrumento a una muestra de estudiantes de la Universidad Autónoma de Baja California (UABC); demuestra que la totalidad de los reactivos evalúan un mismo constructo a partir del análisis de su dimensionalidad ($n=128,791$) y registra una estructura factorial compuesta por dos factores: planeación y gestión del proceso enseñanza-aprendizaje, y conducción y valoración del proceso enseñanza-aprendizaje. La solución factorial logró explicar 75.02 por ciento de la varianza total; se registró un índice de consistencia interna de .98. Mediante el método de validación cruzada con una segunda muestra, la estructura factorial se comprobó con el MEE ($n=462$), con índices de ajuste aceptables (RMSEA=.07; SRMR=.02; CFI=.93 y TLI=.93). Los resultados de este estudio permitieron concluir que el cuestionario integra los componentes de la competencia docente que resultan más significativos para los estudiantes y demostraron su representatividad con las dimensiones del MECD.

El proceso de acumulación de evidencias de validez implica desarrollar estudios que permitan indagar la calidad métrica del instrumento en el tiempo y en poblaciones diversas. Se requiere de instrumentos de medición de la práctica docente que superen las críticas reportadas por especialistas, una de las cuales se relaciona con los escasos estudios sobre confiabilidad y validez, un seguimiento y reportes sobre el estado que guardan las interpretaciones que derivan de su aplicación (Brennan, 2013).

Por lo anterior, con el desarrollo del presente estudio se contestaron las siguientes preguntas: ¿el modelo hipotético-teórico del CECD ajusta a los datos derivados de su aplicación a muestras representativas de la Universidad Autónoma de Chihuahua (UACH) y de la Universidad Autónoma de Yucatán (UADY)?, y ¿las dimensiones del MECD se encuentran representadas en la estructura interna del CECD?

MÉTODO

El presente estudio es una investigación no experimental, de diseño transeccional des-

criptivo (Ortiz, 2016), y la estrategia metodológica que se aplicó para acumular evidencias de validez relativas a la estructura interna fue el análisis factorial confirmatorio, procedimiento propio del MEE.

Muestra y participantes

Se analizó la información de los cuestionarios aplicados a dos muestras representativas de estudiantes de licenciatura. La selección de las unidades de análisis se determinó mediante la fórmula de población finita de Fernández (1998) con un nivel de confianza de 95 por

ciento, misma que fue sobreestimada en un 15 por ciento como medida de reposición de casos por si se obtenían cuestionarios inconclusos. Las muestras quedaron conformadas por 511 estudiantes de la UACH y 503 de la UADY.

En la Tabla 5 se presentan el sexo y edad de los participantes: 55 por ciento hombres y 45 por ciento mujeres en el caso de la UACH, y 46.3 por ciento hombres y 53.7 por ciento mujeres en el caso de la UADY; \bar{x} =21.09 años para los estudiantes de la UACH y \bar{x} =20.47 años para los de la UADY.

Tabla 5. Características de los alumnos participantes

	VARIABLES	UACH (%)	UADY (%)
Sexo	Femenino	45	53.7
	Masculino	55	46.3
Edad	17	1	.8
	18	8.2	6.4
	19	20	27.3
	20	22.3	26.5
	21	14.9	17.3
	22	17.4	10.7
	23	6.8	4.6
	24	2.7	1.8
	25	1.8	2.4
	26 en adelante	4.4	2.2

Nota: UACH: n=511; UADY: n=503.

Fuente: elaboración propia.

Procedimiento

La recolección de los datos en ambas instituciones educativas se realizó bajo el mismo procedimiento. La aplicación fue con lápiz y papel y la integración y análisis de datos se realizó con ayuda del programa de análisis de datos Statistical Package for the Social Sciences (SPSS por sus siglas en inglés) versión 17.0. Además del CECD se incluyó una sección de datos generales con preguntas sobre sexo, edad, características socioeconómicas y valoración del desempeño global del docente y del

propio estudiante. Las actividades relacionadas con la aplicación fueron: a) preparación del formato del cuestionario; b) capacitación de los aplicadores del CECD; c) aplicación del instrumento; y d) integración y depuración de la base de datos.

Análisis de los datos

Mediante el programa de cómputo para el análisis de datos Stata se realizó la modelización confirmatoria con los pasos sugeridos por Kline (2016), los cuales incluyeron: especificación del

modelo; identificación del modelo; selección de medidas, recopilación, preparación y selección de los datos; evaluación del modelo; reespecificación del modelo; informe de resultados.

RESULTADOS

Se reportan las evidencias que determinan si el modelo propuesto de 25 reactivos del CECD (derivado del estudio de Luna y Reyes, 2015) representa a la población de la UACH y UADY, y se confirma si las dimensiones que conforma el MECD se ven reflejadas en la estructura del CECD al documentar los pasos en la modelización de ecuaciones estructurales.

Muestra representativa de la UACH

Los resultados de los pasos seguidos para obtener evidencias de validez de la estructura interna fueron:

1. *Especificación del modelo.* Se dibujó el modelo sometido a prueba, el cual se encuentra constituido por 25 reactivos agrupados en dos variables latentes.

2. *Identificación del modelo y estimación de los parámetros del sistema de ecuaciones estructurales.* Se estimó bajo métodos de máxima verosimilitud y se optó por el análisis estándar propuesto en el programa. La totalidad de las cargas de los factores hacia los indicadores y entre factores muestran valores positivos, que era lo esperado. Los resultados de esta fase

sugieren que la estimación inicial del modelo sobre los coeficientes derivados presentó congruencia con el modelo CECD. Asimismo, se identificó el sentido positivo de las relaciones y correlaciones con la muestra de la UACH. Los grados de libertad resultaron mayores a cero.

3. *Selección de medidas, recopilar, preparar y seleccionar los datos.* La información de este paso se describió en el método. Las medidas previas de confiabilidad y validez se reportan en el apartado “Cuestionario de evaluación de la competencia docente con base en la opinión de los estudiantes”.

4. *Evaluación de los criterios de calidad de ajuste.* Los índices de bondad de ajuste del modelo del CECD de 25 reactivos reflejaron parsimonia sin afectar sustancialmente el ajuste del modelo de medición. Los valores obtenidos en los diferentes indicadores revelan un buen ajuste bajo la estrategia de ajuste aproximado. Los índices de bondad de ajuste chi cuadrada normada, RMSEA, SRMR, CFI y TLI resultaron con valores aceptables, mientras que el índice de ajuste absoluto chi cuadrado no cumplió con la expectativa y la búsqueda de su ajuste a partir de la reespecificación del modelo ponía en riesgo la representatividad de los principios de la fundamentación teórica que subyace en el cuestionario. Se logró obtener la evidencia suficiente de acuerdo con la regla compuesta de Hu y Bentler (1999) con los índices RMSEA y SRMR (Tabla 6).

Tabla 6. Indicadores de ajuste del modelo de medición con la muestra de la UACH

Índice	Criterio límite	UACH
Sig. de χ^2	> 0.50	.00
χ^2	Valores pequeños	930.82
DI	Valores pequeños	271
χ^2 normado	2.0-5.0	3.43
RMSEA	\leq .08	.06
CFI	Cercano a .95	.91
TLI	>.90	.90
SRMR	< 0.05	.04

Nota: UACH n=511.

Fuente: elaboración propia.

Muestra representativa de la UADY

Los resultados relacionados con los tres primeros pasos de la aplicación de esta técnica analítica fueron similares a los encontrados en la muestra de la UACH. En lo relativo a la *evaluación del modelo*, los valores obtenidos de los diferentes índices de bondad de ajuste RMSEA y SRMR se ubicaron en el límite crítico y el caso de chi cuadrada normado en el rango aceptable. La chi cuadrada, CFI y TLI no registraron valores que confirmen su ajuste, por lo tanto, se optó por realizar la fase de modificación del modelo.

En cuanto a la *re-especificación del modelo*, se utilizó la opción de índices de modificación que ofrece el programa para el análisis de datos Stata como recurso para mejorar el ajuste de los datos del modelo del CECD. En la Tabla 7 se muestran el modelo inicial y el modelo con mejor ajuste posterior a la estrategia de modificación. Los valores obtenidos para un ajuste aproximado, chi cuadrada normado, RMSEA y SRMR resultan en un rango adecuado y los valores en los índices CFI y TLI se ubican en límites aceptables. Al igual que en el caso anterior, el modelo se validó bajo la regla propuesta de Hu y Bentler (1999) con los índices RMSEA y SRMR.

Tabla 7. Indicadores de ajuste de los modelos de medición con muestras de la UADY

Índice	Criterio límite	Inicial (25 reactivos)	Re-especificado (21 reactivos)
Sig. de χ^2	> 0.50	.00	.00
χ^2	Valores pequeños	1,257.14	572.90
Df	Valores pequeños	271	184
χ^2 normado	2.0-5.0	4.63	3.11
RMSEA	$\leq .08$.08	.06
CFI	Cercano a .95	.83	.91
TLI	>.90	.81	.90
SRMR	< 0.05	.05	.04

Nota: n=503.

Fuente: elaboración propia.

Como resultado de estas modificaciones se decidió eliminar cuatro ítems del modelo del CECD, dos ítems de la dimensión “planeación y gestión del proceso de enseñanza-aprendizaje” (ítems 3 y 9) y dos de la dimensión “conducción y valoración del proceso de enseñanza-aprendizaje” (ítems 13 y 18).

Con lo anterior, se responde la pregunta inicial al confirmar el modelo del CECD en las muestras representativas de la UACH y UADY, con la regla propuesta de Hu y Bentler (1999). Asimismo, en ambas universidades la estructura del modelo del CECD refleja las tres dimensiones del MECD al contar con ítems representativos de cada una de estas

dimensiones. Además, se registraron índices de consistencia interna del orden de .95 en la UACH y de .92 en la UADY, valores que indican alta confiabilidad.

DISCUSIÓN

Este trabajo parte de reconocer la necesidad de contar con instrumentos de evaluación de la competencia docente válidos, confiables y pertinentes al contexto educativo. Los cuestionarios de evaluación de la docencia con base en la opinión de los estudiantes son la estrategia de evaluación más utilizada en el ámbito universitario, tanto con fines de control

administrativo como para la mejora, por lo tanto, es indispensable que cumplan con los requisitos básicos de calidad de todo instrumento de evaluación.

A pesar de que los cuestionarios de evaluación de la docencia por los estudiantes tienen una larga tradición y uso generalizado, la investigación y diseño de cuestionarios para evaluar la modalidad de enseñanza por competencias es incipiente. A esta situación contribuye el hecho de que el desarrollo de los cuestionarios en muchos casos es incompleto, ya sea porque carecen de fundamento teórico o de acumulación de evidencias empíricas suficientes (Van Der Schaaf *et al.*, 2019). A esta circunstancia se añade que, en el ámbito universitario, la implementación de los planes de estudio por competencias es relativamente reciente (OCDE, 2002; Gonzáles y Wawenaar, 2003).

Al incluir los principios de la modalidad de enseñanza por competencias en la elaboración de los instrumentos de evaluación docente se busca congruencia entre la evaluación y las demandas institucionales. Al mismo tiempo, la utilización de metodologías robustas para la acumulación de evidencias de validez de la estructura interna abona a la validación teórica del CECD. De esta manera, los puntajes derivados de los cuestionarios representan información válida para la retroalimentación de la práctica docente y la toma de decisiones. Acorde con esta perspectiva, el presente trabajo describe los procesos analíticos a los que se ha sido sometido el CECD con el fin de contar con un instrumento confiable y válido; en particular, muestra el uso de la estrategia de modelización confirmatoria con el Modelamiento de Ecuaciones Estructurales (MEE) como una metodología que permite abonar a la acumulación de evidencias de validez basadas en la estructura interna del CECD con muestras de estudiantes de dos universidades públicas mexicanas.

Cabe señalar que el uso del MEE se caracteriza por su utilidad en la explicación de fenómenos complejos ya que, a partir de

una revisión teórica exhaustiva, el investigador construye y somete a prueba un modelo teórico-hipotético.

Asimismo, en relación con los objetivos planteados sobre confirmar el modelo hipotético-deductivo del CECD, éste se ajusta a los datos derivadores de la aplicación a muestras de estudiantes de la UADY y UACH y permitió comprobar que las dimensiones del MECD se encuentran representadas en la estructura interna del CECD. De lo anterior se concluye lo siguiente: se confirmó la estructura factorial del CECD en aplicaciones a muestras de estudiantes de la UADY y UACH; en el caso de la muestra de la UACH, el modelo teórico de 25 reactivos resultó con un ajuste adecuado y, con ello, la representación de las tres dimensiones del MECD y una estabilidad de su estructura bifactorial; mientras que, en el caso de la UADY, se recurrió a la estrategia de modificación del modelo para mejorar las medidas. Como resultado se propone un modelo que conserva la estructura de dos factores, con 21 reactivos que representan las dimensiones del MECD. El ajuste de estos modelos se encuentra sustentado en las estrategias de Hu y Bentler (1999), y Kim y Bentler (2006).

Los fundamentos teóricos y conceptuales que subyacen al CECD sustentaron el modelo teórico al que se ajustaron los datos de las dos muestras, y representan una configuración posible entre otras que también pueden resultar con ajuste, dado que el investigador va construyendo a partir de decisiones fundamentadas los índices de bondad de ajuste y la teoría (Batista y Coenders, 2012; Kline, 2016).

Buscar el ajuste de los datos con base en el tradicional ajuste absoluto del estadígrafo χ^2 que, se sabe, es sensible al tamaño de muestra, representa en ocasiones una vía que pone en riesgo la representatividad de los principios de la teoría de la enseñanza que subyace en el cuestionario, y también es foco de alarma en este tipo de instrumentos (Byrne y Van de Vijver, 2010); por ello, el ajuste de los datos a partir de la estrategia de ajuste aproximado es

una opción viable en estos casos (Hair *et al.*, 2010; Kim y Bentler, 2006). De esta forma, las decisiones sobre qué modificaciones son más adecuadas en el modelo teórico propuesto recaen en el investigador, que debe velar porque la teoría se vea reflejada en la estructura interna del instrumento sin que los valores de los índices de ajuste se vean afectados.

Por lo anterior, la contribución de este estudio es brindar un instrumento que cuenta con propiedades psicométricas adecuadas y con una estructura interna que responde a los principios de la teoría que le subyace: el enfoque de enseñanza por competencias. En este sentido, representa un instrumento válido y congruente con la pedagogía que se le demanda al docente en las dos universidades públicas donde se confirmó el CECD.

Dado que la estructura interna del CECD se confirmó en las dos universidades (UACH y UADY), que se añaden el caso de la UABC (Luna y Reyes, 2015), es factible concluir que se dispone de un cuestionario válido y confiable para tres universidades públicas de México ubicadas en las regiones norte, sureste y noroeste del país.

La revisión de la literatura especializada que da cuenta del diseño, desarrollo y validación de cuestionarios orientados a evaluar la competencia docente permitió observar la

carencia de instrumentos de medida que descansen en un andamiaje teórico sólido (Van Der Schaaf *et al.*, 2019; Luna, 2010). Estos instrumentos no sólo suelen obviar procesos de adaptación al contexto educativo mexicano, también omiten el empleo de procedimientos orientados a asegurar su equivalencia métrica que respalde la comparación entre estudiantes pertenecientes a distintos grupos de referencia. Aplicaciones futuras del CECD deberán contribuir a solventar dichas carencias mediante el empleo de procedimientos de invarianza factorial que aseguren la equivalencia de la medida como resultado de la aplicación del mismo instrumento a muestras pertenecientes a distintas poblaciones (Gunnell *et al.*, 2012).

Finalmente, si bien la estructura interna del CECD reproduce la configuración del MEE en las muestras de estudiantes de las tres instituciones participantes (las diferencias en la cantidad de reactivos es mínima), se reconoce que contar con un instrumento válido es una condición necesaria, pero no suficiente, para acceder a la complejidad de la práctica docente. Lo anterior refuerza la necesidad de utilizar diversos mecanismos y estrategias que permitan una evaluación integral de la actividad docente que reconozca, como punto de partida, su naturaleza y complejidad, y como propósito, mejorarla.

REFERENCIAS

- American Educational Research Association/American Psychological Association/National Council on Measurement in Education (2014), *Standards for Educational and Psychological Testing*, Washington, DC, American Educational Research Association.
- ANDRADE-Abarca, Paola, Lizeth Ramón-Jaramillo y María Loaiza-Aguirre (2018), "Aplicación del SEEQ como instrumento para evaluar la actividad docente universitaria", *Revista de Investigación Educativa*, vol. 26, núm. 1, pp. 259-275. DOI: <http://doi.org/10.6018/rie.36.1.260741>
- ARBUCKLE, James (2013), *IBM SPSS Amos 22 User's Guide*, Chicago, Amos Development Corporation.
- BATISTA, Joan y Gema Coenders (2012), *Modelos de ecuaciones estructurales*, Madrid, La Muralla.
- BENTON, Stephen y Dan Li (2015), "Validity and Reliability of IDEA, "Teaching Essentials", IDEA, en: <https://eric.ed.gov/?id=ED573675> (consulta: 8 de diciembre de 2019).
- BENTON, Stephen, Dan Li, Ron Brown, Meixi Guo y Patricia Sullivan (2015), "Revising the IDEA Student Ratings of Instruction System: 2002–2011 Data", IDEA, en: https://www.ideaedu.org/Portals/0/Uploads/Documents/Technical-Reports/Revising-the-IDEA-Student-Ratings-of-Instruction-System-2002-2011-Data_Technical_report_18.pdf (consulta: 8 de diciembre de 2019).

- BORSBOOM, Denny (2006), "The Attack of the Psychometricians", *Psychometrika*, vol. 71, núm. 3, pp. 425-440. DOI: <https://doi.org/10.1007/s11336-006-1447-6>
- BOYSEN, Guy (2016), "Using Student Evaluations to Improve Teaching: Evidence-based recommendations", *Scholarship of Teaching and Learning in Psychology*, vol. 2, núm. 4, pp. 273-284. DOI: <https://doi.org/10.1037/stl0000069>
- BRENNAN, Robert (2013), "Commentary on Validating the Interpretations and Uses of Test Scores", *Journal of Educational Measurement Spring*, vol. 50, núm. 1, pp. 74-83.
- BUJANDA, Elena (2017), "Diseño de un cuestionario de competencias docentes del profesorado de Formación Profesional de la Familia, Servicios Socioculturales y a la Comunidad", *Revista de Educación Social*, núm. 24, pp. 1163-1178.
- BYRNE, Barbara y Fons Van de Vijver (2010), "Testing for Measurement and Structural Equivalence in Large-Scale Cross-Cultural Studies: Addressing the issue of nonequivalence", *International Journal of Testing*, vol. 10, núm. 2, pp. 107-132. DOI: <https://doi.org/10.1080/15305051003637306>
- CANO, Elena (2015), "Las rúbricas como instrumentos de evaluación de competencias en educación superior: juro o abuso?", *Profesorado. Revista de Currículum y Formación del Profesorado*, vol. 19, núm. 2, pp. 265-280.
- CISNEROS-Cohernour, Edith, Benilde García-Cabrero, Edna Luna y Rigoberto Marín (2012), *Evaluación de competencias docentes en la educación superior*, México, Red para el Desarrollo y Evaluación de Competencias Académicas/Juan Pablos Editor.
- CÓRDOVA, Alfredo, Julia Moreno, Mihaela Stegaru y Carlos Staff (2015), "Construcción de un instrumento para evaluar competencias profesionales durante la formación preclínica en Medicina", *Investigación en Educación Médica*, vol. 4, núm. 15, pp. 145-154. DOI: <https://doi.org/10.1016/j.riem.2015.01.001>
- DENYER, Monique, Jacques Furnemont, Roger Poulain y Georges Vanloubbeek (2007), *Las competencias en la educación, un balance*, México, Fondo de Cultura Económica.
- FELMAN, Kenneth (2007), "Identifying Exemplary Teachers and Teaching: Evidence from student ratings", en Raymond Perry y John Smart (eds.), *The Scholarship of Teaching and Learning in Higher Education: An evidence-based perspective*, Dordrecht, Springer, pp. 93-129.
- FERNÁNDEZ Nogales, Ángel (1998), *Investigación de mercados: obtención de información*, Madrid, Ed. Civitas, Biblioteca Civitas Economía y Empresa, colección Empresa.
- GARCÍA-Cabrero, Benilde, Javier Loredo y Guadalupe Carranza (2008a), "Análisis de la práctica educativa de los docentes: pensamiento, interacción y reflexión", *Revista Electrónica de Investigación Educativa*, vol. 10, núm. especial, pp. 1-15, en: <http://redie.uabc.mx/NumEsp1/contenido-garcialoredocarranza.html> (consulta: 8 de diciembre de 2019).
- GARCÍA-Cabrero, Benilde, Javier Loredo, Guadalupe Carranza, Alma Elena Figueroa, Isabel Arbesú, Miguel Monroy y Rafael Reyes (2008b), "Las aproximaciones teórico-metodológicas en los trabajos de la RIED: consideraciones en torno a la construcción de un modelo de evaluación de la práctica docente", en Mario Rueda (ed.), *La evaluación de los profesores como recurso para mejorar su práctica*, México, UNAM-IISUE/Plaza y Valdés, pp. 162-220.
- GARCÍA-Cabrero, Benilde, Javier Loredo, Edna Luna y Mario Rueda (2014), *Competencias docentes en educación media y superior. Desarrollo y validación de un modelo de evaluación*, México, UABC/Juan Pablos Editor.
- GARGALLO, Bernardo, Jesús Suárez, Pedro Garfella y Amparo Fernández (2011), "El cuestionario CEMEDEPU. Un instrumento para la evaluación de la metodología docente y evaluativa de los profesores universitarios", *Estudios sobre Educación*, vol. 21, pp. 9-40.
- GONZÁLES, Julia y Robert Wawenaar (eds.) (2003), *Tuning Educational Structures in Europe. Informe final. Fase uno*, Bilbao, Universidad de Deusto, en: http://tuningacademy.org/wp-content/uploads/2014/02/TuningEUI_Final-Report_SP.pdf (consulta: 8 de diciembre de 2019).
- GUARDIA, Joan (2016), "Esquema y recomendaciones para el uso de los Modelos de Ecuaciones Estructurales", *Revista de Estudios e Investigación en Psicología y Educación*, vol. 3, núm. 2, pp. 75-80. DOI: <https://doi.org/10.17979/riep.2016.3.2.1847>
- GUNNELL, Katie, Philip Wilson, Bruno Zumbo, Diane Mack y Peter Crocker (2012), "Assessing Psychological Nedd Satisfacción in Exercise Contexts: Issues of score invariance, item modification, and context", *Measurement in Physical Education and Exercise Science*, vol. 16, núm. 3, pp. 219-236.
- HAIR, Joseph, William Black, Barry Babin y Rolph Anderson (2010), *Multivariate Data Analysis*, New Jersey, Pearson Prentice Hall.
- HOOPER, Daire, Joseph Coughlan y Michael Mullen (2008), "Structural Equation Modelling: Guidelines for determining model fit", *Electronic Journal of Business Research Methods*, vol. 6, núm. 1, pp. 53-60.
- HORNSTEIN, Henry (2017), "Student Evaluations of Teaching are an Inadequate Assessment Tool for Evaluating Faculty Performance", *Cogent Education*, vol. 4, núm. 1. DOI: <https://doi.org/10.1080/2331186X.2017.1304016>
- HOUSE, Ernest (1998), "Acuerdos institucionales para la evaluación", *Perspectivas*, vol. 28, núm. 1, pp. 123-131.

- HU, Li-tze y Peter Bentler (1999), "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional criteria versus new alternatives", *Structural Equation Modeling*, vol. 6, núm. 1, pp. 1-55. DOI: <https://doi.org/10.1080/10705519909540118>
- KIM, Kevin y Peter Bentler (2006), "Data Modeling: Structural equation modeling", en Judith Green, Gregory Camili y Patricia Elmore (eds.), *Handbook of Complementary Methods in Education Research*, Nueva York, American Education Research Association, pp. 161-175. DOI: <https://doi.org/10.4324/9780203874769>
- KLINE, Rex (2016), *Principles and Practice of Structural Equation Modeling*, Nueva York, Guilford Press.
- LINSE, Angela (2017), "Interpreting and Using Student Ratings Data: Guidance for faculty serving as administrators and on evaluation committees", *Studies in Educational Evaluation*, vol. 54, pp. 94-106.
- LUNA, Edna (2010), "Aspectos implícitos sobre la enseñanza reflejados en los cuestionarios de evaluación de la docencia", en Mario Rueda y Frida Díaz-Barriga Arceo (eds.), *Evaluación de la docencia. Perspectivas actuales*, México, Paidós, pp. 63-83.
- LUNA, Edna y Erika Reyes (2015), "Validación de constructo de un cuestionario de evaluación de la competencia docente", *Revista Electrónica de Investigación Educativa*, vol. 17, núm. 3, pp. 13-27, en: <http://redie.uabc.mx/vol17no3/contenido-luna-pinuelas.html> (consulta: 8 de diciembre de 2019).
- MARSH, Herbert (2007), "Students' Evaluations of University Teaching: Dimensionality, reliability, validity, potential biases and usefulness", en Raymond Perry y John Smart (eds.), *The Scholarship of Teaching and Learning in Higher Education: An evidence-based perspective*, Dordrecht, Springer, pp. 319-383.
- MESSICK, Samuel (1995), "Validity of Psychological Assessment, Validation of Inferences from Person's Responses and Performances as Scientific Inquiry into Score Meaning", *American Psychologist*, vol. 50, núm. 9, pp. 741-749.
- MORENO, Tiburcio (2018), "La evaluación docente en la universidad: visiones de los alumnos", *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, vol. 16, núm. 3, pp. 87-101. DOI: <https://doi.org/10.15366/reice2018.16.3.005>
- MUÑIZ, José y Eduardo Fonseca-Pedrero (2008), "Construcción de instrumentos de medida para la evaluación universitaria", *Revista de Investigación en Educación*, vol. 5, pp. 13-25.
- Organisation for Economic Co-operation and Development (OECD) (2002), *Definition and Selection of Competencies: Theoretical and conceptual foundations (DeSeCo) Annual Report*, París, OECD, en: <https://deseco.ch/bfs/deseco/en/index/01.parsys.70925.downloadList.59988.DownloadFile.tmp/2001annualreport.pdf> (consulta: 8 de diciembre de 2019).
- ORTIZ Uribe, Frida (2016), *Diccionario de metodología de la investigación científica*, Madrid, Limusa.
- Red Iberoamericana de Investigadores de la Evaluación de la Docencia (2008), "Reflexiones sobre el diseño y puesta en marcha de programas de evaluación de la docencia", *Revista Iberoamericana de Evaluación Educativa*, vol. 1, núm. 3, pp. 163-168.
- RIOS, Joseph y Craig Wells (2014), "Validity Evidence Based on Internal Structure", *Psicothema*, vol. 26, núm. 1, pp. 108-16. DOI: <https://doi.org/10.7334/psicothema2013.260>
- ROYAL, Kenneth (2017), "A Guide for Making Valid Interpretations of Student Evaluation of Teaching (SET) Results", *Journal of Veterinary Medical Education*, vol. 44, núm. 2, pp. 316-322, en: <https://jvme.utpjournals.press/doi/abs/10.3138/jvme.1215-201R> (consulta: 8 de diciembre de 2019).
- RUÍZ, Miguel, Antonio Pardo y Rafael San Martín (2010), "Modelos de ecuaciones estructurales", *Papeles del Psicólogo*, vol. 31, núm. 1, pp. 34-45.
- SCHREIBER, James, Amaury Nora, Frances Stage, Elizabeth Barlow y Jamie King (2006), "Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A review", *The Journal of Educational Research*, vol. 99, núm. 6, pp. 323-337.
- SPOOREN, Pieter, Bert Brockx y Dimitri Mortelmans (2013), "On the Validity of Student Evaluation of Teaching", *Review of Educational Research*, vol. 83, núm. 4, pp. 598-642. DOI: <https://doi.org/10.3102/0034654313496870>
- VAN DER SCHAAF, Marieke, Bert Slof, Leonie Boven y Angela de Jong (2019), "Evidence for Measuring Teachers' Core Practices", *European Journal of Teacher Education*, vol. 42, núm. 5, pp. 675-694. DOI: <https://doi.org/10.1080/02619768.2019.1652903>