

ANÁLISIS DE VALIDEZ DEL CUESTIONARIO DE EVALUACIÓN DE LA COMPETENCIA DOCENTE (CECD) CON BASE EN LA OPINIÓN DE ALUMNOS UNIVERSITARIOS

ERIKA PAOLA REYES PIÑUELAS/ EDNA LUNA SERRANO
UABC

RESUMEN: Este trabajo amplia los análisis de confiabilidad y validez de constructo del Cuestionario de Evaluación de la Competencia Docente con base en la opinión de estudiantes. Se estudiaron muestras de puntajes de tres universidades públicas. El cuestionario consta de 35 reactivos y su estructura teórica responde al Modelo de Evaluación de Competencias Docentes. En el Estudio 1 con muestras de la Universidad Autónoma de Baja California (UABC) se replicaron los análisis de unidimensionalidad ($n=128,791$) y de AFE ($n=67,111$). El resultado del análisis de unidimensionalidad demostró que 32 reactivos evalúan el mismo constructo; el AFE reveló dos factores: Planeación y gestión del proceso enseñanza-aprendizaje, y, Conducción y valoración del proceso enseñanza-aprendizaje, con una varianza total de 75.02% y un alfa de Cronbach de .98. En el Estudio 2 se comprobó la estructura producto del AFE

con modelamiento de ecuaciones estructurales con muestras de la UABC ($n=462$), Universidad Autónoma de Chihuahua (UACH) ($n=511$) y Universidad Autónoma de Yucatán (UADY) ($n=503$). Los resultados de la UABC y la UACH ajustaron a un modelo alternativo bifactorial conformado por 25 reactivos; los valores en índices de bondad de ajuste fueron: en RMSEA de .79 en UABC y .69 en UACH y un CFI con .937 y .914 respectivamente. La muestra de la UADY posterior al ajuste del modelo alternativo quedó conformado por una estructura bifactorial con 21 reactivos, RMSEA=.065 y CFI=.91. Los resultados fundamentan la realización de inferencias válidas y confiables y respaldan su utilización en estas universidades.

PALABRAS CLAVE: validez de las pruebas, evaluación de profesores, encuestas, competencias docentes, análisis factorial.

Introducción

Los cuestionarios con base en la opinión de los alumnos son la técnica de recopilación de información del quehacer docente de mayor presencia tanto en el ámbito internacional como nacional. Su uso masivo se facilita por el corto tiempo requerido para su aplicación y bajo costo.

Si bien, se reconoce que dichos cuestionarios tienden a ser estadísticamente confiables y válidos, la problemática en torno a su uso se concentra en que, promueven una pedagogía conservadora (Centra, 1993) al no ser contruidos en congruencia con los modelos educativos actuales, asimismo, que se apegan de manera limitada a las teoría de enseñanza que les subyace (Luna, 2000), y se cuestiona su uso indiscriminado sin atender los aspectos básicos de confiabilidad y validez (Berk, 2006).

En consideración a lo anterior, el presente trabajo tiene el propósito de ampliar las evidencias de validez de constructo del Cuestionario de Evaluación de la Competencia Docente (CECD) bajo una rigurosidad metodológica con muestras de estudiantes distintas a la población en la que fue originalmente validado. De esta forma se fundamenta su uso en otros contextos universitarios.

En un estudio previo Luna, Calderón, Caso y Cordero (2012) reportaron el proceso de construcción y validación del cuestionario del CECD desarrollado en congruencia con el modelo de enseñanza por competencias. Dado que en este estudio se presentaron condiciones diferentes (tamaño de muestra y aplicación en papel y lápiz) se consideró pertinente identificar la estructura factorial con propósitos exploratorios y obtener suficientes conocimientos para formular hipótesis concretas sobre la relación entre indicadores y dimensiones latentes con una muestra que mantuviera las mismas condiciones de aplicación en los análisis de datos.

Consideraciones teóricas

La base teórica que sustenta al CECD es el Modelo de Evaluación de Competencias Docentes (MECD) para la educación media y superior de García-Cabrero, Loredó, Luna y Rueda (2008), este plantea tres momentos o dimensiones presentes para el desarrollo de la función docente.

1. Previsión del proceso enseñanza-aprendizaje, son las actividades previas incluye el pensamiento del docente.
2. Conducción del proceso enseñanza-aprendizaje, relacionado con la interacción del docente y alumno en el aula.

3. Valoración de proceso enseñanza-aprendizaje, es la reflexión en torno a los resultados alcanzados por el ejercicio de la práctica docente y evaluadas por diferentes actores como alumnos, docentes, los pares y directivos.

En específico el CECD consta de 35 reactivos tipo Likert con cinco niveles de respuesta, al cual le precede un estudio sobre el desarrollo y validación reportado por Luna et al. (2012), en suma los resultados arrojaron: una confiabilidad con la prueba Alfa de Cronbach de .967; en el análisis de unidimensionalidad se concluyó que los reactivos se comportaron según las expectativas (excepto los reactivos 34 y 35); en el AFE resultaron tres dimensiones: *Planeación y gestión del proceso enseñanza-aprendizaje*, *Interacción didáctica en el aula*, *Evaluación y comunicación del proceso de enseñanza-aprendizaje*, y como indicador, *Tecnologías de la información y la comunicación*, que explican un 59.3% de varianza. Se concluyó que el cuestionario cuenta con evidencias de validez y confiabilidad para evaluar el constructo de competencia docente.

Método

Con el propósito de ampliar la información sobre la validez y confiabilidad, se realizaron dos estudios: Estudio 1, se enfocó en confirmar que los reactivos de dicho cuestionario midieran el mismo rasgo latente y conocer con propósitos exploratorios la estructura factorial con datos de la población de la UABC; Estudio 2, con propósitos confirmatorios se dirigió a corroborar la estructura factorial del CECD en muestras de alumnos de la UABC, y confirmar la generalización de los resultados en las muestras de la UACH y UADY.

Estudio 1

Muestra y participantes. Los cuestionarios analizados involucraron a 37,094 estudiantes de licenciatura de la UABC adscritos a los tres campus, en el segundo periodo del 2011. El total de cuestionarios disponibles para analizar fue mayor a 37,094, ya que los alumnos contestaron un cuestionario por cada asignatura cursada en el semestre. Cabe señalar que la información general relacionada con periodo de aplicación, nivel de licenciatura, lugar de adscripción de los estudiantes (caracterización de la muestra) se obtiene en cada muestra utilizada de la UABC.

En el análisis de unidimensionalidad la muestra quedó constituida por 128,791 cuestionarios. Para el resto de los análisis de la muestra UABC, se utilizó el criterio de replicabilidad por medio del método de validación cruzada, por consiguiente se obtuvieron

dos muestras aleatorias (con ayuda del SPSS versión 17.0). La submuestra A conformada por 66,148 cuestionarios utilizada para el AFE y la submuestra B (67,111) para el AFC esta última forma parte del Estudio 2. Las características de la muestra se describen en la tabla 1.

Procedimiento. Los puntajes de los cuestionarios se obtuvieron de la base de datos institucional a través del Sistema de Evaluación Docente de la UABC y se generaron a partir de respuestas de los estudiantes al completar el CECD por medio de un portal en internet. Se replicaron los análisis de unidimensionalidad y AFE reportados por Luna et al. (2012) mismos que a continuación se describen:

1. Análisis de unidimensionalidad del modelo Rasch Masters: con el apoyo del programa de análisis de datos Winsteps. Se obtuvieron estadígrafos tales como: correlación punto biserial, discriminación del reactivo e índices de ajuste.
2. AFE: se realizó con programa SPSS (versión 17.0), por medio del método de extracción de componentes principales, con las rotaciones, ortogonal (varimax) y oblicua (promax), dado que es recomendable hacer ambas rotaciones pues si la agrupación factorial es idéntica en ambas entonces la estructura factorial del instrumento es clara y estable. Se valoraron los factores susceptibles de ser interpretados a través de diferentes criterios de decisión como el de raíz latente y contraste de caída, prueba de comunalidades, varianza explicada y saturaciones de los factores.

En cada uno de los análisis realizados se estimó el coeficiente Alfa de Cronbach del instrumento en general.

Resultados

Análisis de unidimensionalidad. Se realizó este análisis a 32 reactivos y se obtuvo una consistencia interna de .97 con alfa de Cronbach. Los resultados para la determinación de la unidimensionalidad del test tales como: correlaciones punto biserial, discriminación y dificultad de los reactivos y valores de bondad de ajuste próximo (*infit*) y lejano (*oufit*) se

comportaron de acuerdo a la expectativa. Con lo cual se demostró que los 32 reactivos miden un mismo constructo.

Cabe señalar que se excluyeron de la muestra y análisis anterior los reactivos 1, 34 y 35 (por presentar valores fuera de los umbrales críticos) y cuestionarios con puntuaciones extremas relacionadas con una alta y baja competencia docente (con el propósito reducir la influencia de los valores atípicos), dichos resultados obtenidos a partir de un análisis de unidimensionalidad previo dentro del Estudio 1.

Análisis Factorial Exploratorio. Se agruparon los 32 reactivos en dos factores interpretables. Los criterios de raíz latente y criterio de contraste de caída permitieron justificar dos factores a considerarse en la interpretación. Para la determinación de los factores en la prueba de comunalidades se cubrió un mínimo de .50 para ser considerado dentro de un factor. La mayoría de las cargas factoriales mostraron saturaciones altas en uno de los factores y cargas bajas o insignificantes en el otro y se cumplió la expectativa de al menos tres reactivos por factor.

Los pesos y correlaciones obtenidos con la rotación varimax y promax presentaron un comportamiento idéntico al arrojar dos factores y obtener una varianza explicada en el Factor 1 de 71.43%, en el Factor 2 de 3.59% con un total de 75.02%, superando el 60% para las ciencias sociales (según Hair et al., 1999). Posteriormente, se renombraron los dos factores de acuerdo a la concordancia en una o más dimensiones del modelo teórico, así el conjunto de reactivos del Factor 1 coincide con las dimensiones de *Conducción y valoración del proceso enseñanza-aprendizaje* por tanto así fue nombrado y los reactivos del Factor 2, a las de *Planeación y gestión de proceso enseñanza-aprendizaje*. La prueba estadística alfa de Cronbach resultó con .987 en el total de los reactivos (Factor 1=.981; Factor 2=.972). Dado los resultados satisfactorios en este análisis se procedió al siguiente estudio.

Estudio 2

Muestras y participantes. Se analizaron los datos de muestras de estudiantes de licenciatura que cursaban el segundo periodo del 2011.

En la UABC el análisis se realizó con la submuestra B (67,111). En la UACH los cuestionarios analizados involucraron a estudiantes adscritos a unidades académicas de la ciudad de Chihuahua, Chih. La muestra aleatoria se determinó por la fórmula según Scheaffer y Mendelhall (1987) resultó de 393, obtenida a partir de una población de 21,978

estudiantes, se agregó un 15% más como medida en caso de obtener cuestionarios incompletos, la muestra quedó conformada por 511, ésta posteriormente se distribuyó según el tamaño de la población por área del conocimiento.

En la UADY los cuestionarios analizados involucraron a 503 estudiantes adscritos a unidades académicas de la ciudad de Mérida, Yucatán. Esta muestra aleatoria se obtuvo igual a la obtenida por la UACH, en este caso a partir de una población de 11,997 estudiantes. La información sobre las características de las tres muestras de los participantes se puede revisar en la tabla 1.

Tabla 1.

Porcentajes de las características de los participantes

Variables		UABC	UACH	UADY
Sexo	Femenino	53.7	45	53.7
	Masculino	43.3	55	46.3
Edad	17	0.0	1.0	.8
	18	6.6	8.2	6.4
	19	17.8	20.0	27.3
	20	18.6	22.3	26.5
	21	17.3	14.9	17.3
	22	12.9	17.4	10.7
	23	8.3	6.8	4.6
	24	5.0	2.7	1.8
	25	3.3	1.8	2.4
	26 en adelante	10.2	4.4	2.2

Nota: UABC: n=37,094. UACH: n=511. UADY: n=503.

Procedimiento. La muestra de la UABC se obtuvo a través de una aplicación en línea. Para los casos de la UACH y la UADY la aplicación fue de lápiz y papel. Se integró una base de datos para cada muestra con el programa de análisis de datos SPSS versión 17.0, mismas que se analizaron a través del programa Stata (StataCorp., 2011). El análisis de los datos

se realizó con el AFC mediante la metodología del modelamiento de ecuaciones estructurales. Para evaluar el modelo se utilizaron seis índices de bondad de ajuste: chi cuadrada (χ^2), chi cuadrada normado, Raíz del residuo cuadrático promedio de aproximación (RMSEA), Raíz del residuo cuadrático promedio estandarizado (SRMR), Índice de bondad de ajuste comparativo (CFI) e Índice de Tucker-Lewis (TLI). El estadígrafo de χ^2 es la medida tradicional para evaluar el ajuste global del modelo propuesto, sin embargo Hopper *et al.* (2008) reportaron que presenta limitaciones entre las que destacan: que la prueba de la significación estadística de la χ^2 es sensible a muestras grandes, lo cual resulta con frecuencia en rechazo del modelo; y en el caso de muestras pequeñas presenta dificultades para discriminar entre un buen ajuste de otro con ajuste pobre. Complementar la valoración del modelo con otros índices ha sido la alternativa, por ejemplo dos índices que no presentan estas limitaciones son chi cuadrado normado que también es un índice de ajuste global y este tiene la cualidad de reducir al mínimo el impacto del tamaño de la muestra, así como RMSEA (según Hopper et al.), de esta forma la estrategia adecuada fue incluir varios índices en la valoración del modelo.

Resultados

UABC. Se confirmó la estructura factorial del CECD resultado del AFE con la submuestra B de 67,111 cuestionarios. El análisis se estimó con el método de máxima verosimilitud. Los resultados de los seis índices de bondad de ajuste utilizados para valorar el modelo cuatro arrojaron valores aceptables, con excepción de χ^2 (203,830.17, con 463 gl) y chi cuadrada normado (440.23), en el caso de χ^2 el resultado implica que no hay ajuste entre el modelo y los datos, por tanto se realizaron dos estrategias que redundaron en una mejora del modelo:

1. Disminuir el tamaño de la muestra, ya que, χ^2 presenta sensibilidad al tamaño de muestra al igual que CFI (Hooper et al., 2008), por lo tanto, se optó por una muestra aleatoria de 462 cuestionarios.
2. Revisar los recursos de la información de salida del programa Stata (en la opción de Índices de Modificación) el cual se presentan posibles relaciones en el modelo original que pueden ser incluirlas en la re-especificación.

Así, el modelo alterno del CECD quedó conformado por dos factores con 25 reactivos. El procedimiento de re-especificación del modelo sugirió prescindir de los

reactivos 3, 11 y 13 del Factor 1, y los reactivos 6, 18, 21 y 29 del Factor 2. Asimismo, de forma adicional para mejorar el ajuste de los datos se correlacionaron los errores de reactivos 2-4, 30-31 y 32-33.

Como resultado de la re-especificación del modelo, disminuyó el valor de χ^2 a 1,050.72 con 271 gl, un SRMR con una variación dentro de la expectativa y el resto de los índices de ajuste mejoraron, en la tabla 2 se muestran los valores de este modelo. El índice de ajuste global χ^2 indica que el modelo posterior a la re-especificación continúa sin ajustar a los datos, sin embargo, el resto de los indicadores muestran un ajuste del modelo aceptable. Si bien los índices son una guía útil para el proceso de validación, también es cierto que en este proceso la teoría que subyace al instrumento debe ser revisada en cada paso. El modelo alterno presenta una agrupación y relaciones congruentes con el MECD. Por lo tanto los resultados aceptables en los índices de ajuste a la par de una cuidadosa revisión de la teoría que le subyace en el proceso de análisis, permiten afirmar que las inferencias a partir de los resultados de la aplicación del CECD son válidas y confiables.

Posteriormente se confirmó la generalización del CECD en las muestras de la UADY y UACH, la tabla 2 muestra los valores de los índices de ajuste examinados.

Tabla 2.

Indicadores de ajuste de los modelos de medición

Índice	Criterio límite	UABC* n=462	UACH* n=511	UADY** n=503
χ^2	> 0.50	.00	.00	.00
χ^2 normado	2.0-5.0	3.87	3.43	3.11
RMSEA	\leq .08	.079	.069	.065
CFI	Cercano	.937	.914	.916

	a .95			
TLI	> .90	.931	.904	.905
SRMR	< 0.05	.029	.041	.047

Nota: *Índices de ajuste con modelo de 25 reactivos. **Índices de ajuste con 21 reactivos.

UACH. Los índices de bondad de ajuste del modelo alternativo del CECD con la muestra de esta universidad resultaron aceptables, a excepción de χ^2 (930.82, con 271 gl) que se ubicó fuera de los límites críticos, por lo tanto, bajo los argumentos planteados en el análisis precedente de la muestra de la UABC se confirma la validez y confiabilidad de las puntuaciones a partir de la aplicación del CECD para el caso UACH.

UADY. Los valores obtenidos de los índices de bondad de ajuste del modelo alternativo del CECD resultaron fuera de la expectativa, excepto el índice de ajuste chi cuadrada normado, por lo tanto, se procedió a la modificación del modelo alternativo a partir de la estrategia conjunta entre los recursos de información de salida del programa Stata y la revisión de la teoría que subyace en el instrumento. El modelo que ajustó a los datos quedó conformado por 21 reactivos, los índices de ajuste presentan valores aceptables, a excepción de la χ^2 (1,050.72, con 271 dl). Dicho modelo presentó una agrupación y relaciones congruentes con el MECD, por lo cual los resultados a partir de su aplicación son válidos y confiables para dicha universidad.

Conclusiones

Las fortalezas metodológicas del CECD residen en las evidencias de validez y confiabilidad reportadas en el presente trabajo que se suman a las obtenidas en el estudio de Luna et al. (2012). Los resultados indican valores adecuados en los criterios para la determinación de la unidimensionalidad del test, estabilidad en la estructura factorial, un porcentaje de varianza explicada que supera la considerada para las ciencias sociales, donde se reflejan las tres dimensiones del MECD, índices altos de confiabilidad y en el caso de los índices bondad de ajuste en los modelos alternos de las muestras de la UABC, UACH y UADY valores aceptables, con excepción de χ^2 , dicho resultado muestra la relevancia evaluar el

modelo tomando en cuenta más de un índice de ajuste y en consideración con la teoría que subyace al instrumento, estrategia planteada por Hooper et al. (2008).

El proceso de validación del CECD ha permitido constatar que el modelo que subyace al instrumento ajusta de manera aceptable a los datos, se pueden hacer inferencias validas y confiables a partir de su aplicación, y tiene elementos necesarios para fundamentar su utilización en otras universidades. De esta forma, se constituye como un instrumento de evaluación de la práctica docente en la modalidad por competencias que cuenta con suficientes evidencias de validez y confiabilidad.

Bibliografía

- Berk, R. (2006). *Thirteen Strategies to Measure College Teaching: a consumer's guide to rating scale construction, assessment, and decision making for faculty, administrators, and clinicians*. Virginia, EE.UU.: Stylus Publishing.
- Centra, J. (1993). *Reflective faculty evaluation: enhancing teaching and determining faculty effectiveness*. Nueva York, EE.UU.: Jossey-Bass.
- García, B., Loredó, J., Luna, E., y Rueda, M. (2008). Modelo de Evaluación de Competencias Docentes para la Educación Media y Superior. *Revista Iberoamericana de Evaluación Educativa 2008*, 1(3e), 96-108. Recuperado de http://www.rinace.net/riee/numeros/vol1-num3_e/art8.html
- Hair, J., Anderson, R., Tatham, R. y Black, W. (1999). *Análisis Multivariante*. Madrid, España: Pearson/Prentice Hall.
- Hooper, D., Coughlan, J. & Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, 6, 1, 53-60.
- Luna, E. (2000). Aspectos implícitos sobre la enseñanza reflejados en los cuestionarios de evaluación de la docencia. En M. Rueda y F. Díaz-Barriga (Comps.) *Evaluación de la docencia. Perspectivas actuales* (pp. 63-84). México: Paidós Educador.
- Luna, E., Calderón, N., Caso, J. y Cordero, G. (2012). Desarrollo y validación de un cuestionario de evaluación de la competencia docente con base en la opinión de los estudiantes. En E. Cisneros-Cohernour, B. García-Cabrero, E. Luna y R. Marín (coords.), *Evaluación de Competencias Docentes en la Educación Superior* (pp.119-158). México: Red para el Desarrollo y Evaluación de Competencias Académicas; Juan Pablos Editor.