Diseño y validación de una medida de práctica docente para educación superior

José Ángel Vera Noriega, Claudia Karina Rodríguez Carvajal, Francisca Leticia Medina Figueroa y Lorena Patricia Gerardo Muñoz

Resumen

En este estudio se describen las propiedades psicométricas de un instrumento para la evaluación docente en nivel superior, con una metodología de corte cuantitativo. Se encontró la estructura factorial de la medida y se comparó con la propuesta teórica a través de un análisis factorial el cual arrojó cuatro factores: *a)* Didáctica, *b)* Evaluación, *c)* Planeación y *d)* Motivación, con un total de 54 reactivos y una varianza explicada de 55.53%.

Abstract

This study describes the psychometric properties of an instrument for teacher evaluation at undergraduate level, with a methodology of quantitative nature. We found the factor structure of the measure and it was compared to the theoretical proposal through a factor analysis which yielded four factors: a) teaching, b) evaluation, c) planning, and d) Motivation, with a total of 54 questions and an explained variance of 55.53%.

José Ángel Vera Noriega. Centro de Investigación en Alimentación y Desarrollo A.C. Departamento de Desarrollo Humano y Bienestar Social. [avera@ciad.mx]

Claudia Karina Rodríguez Carvajal. Centro de Investigación en Alimentación y Desarrollo A.C. Departamento de Desarrollo Humano y Bienestar Social. [psicología@ciad.mx]

Francisca Leticia Medina Figueroa. Instituto de Formación Docente del Estado de Sonora. Coordinación de Formación Inicial. [lemefi63@hotmail.com]

Lorena Patricia Gerardo Muñoz. Instituto de Formación Docente del Estado de Sonora. Coordinación de Formación Inicial. [p.gerardo@ifodes.edu.mx]

Revista Intercontinental de Psicología y Educación, vol. 14, núm. 2, julio-diciembre 2012, pp. 165-177.
Fecha de recepción: 27 de octubre de 2011 | fecha de aceptación: 17 de febrero de 2012.

Anteriormente, se llevó a cabo la validación por contenido la cual involucró la participación de 20 directivos y profesores en la revisión de los reactivos con el objeto de que se señalaran omisiones, errores sintácticos y semánticos. La confiabilidad del instrumento fue de .96, lo que se considera aceptable.

Palabras clave
evaluación docente, validez, confiabilidad, educación superior.

Previously, a content validation was conducted, which involved the participation of 20 managers and teachers in the review of the questions with the aim of highlighting omissions, syntactic and semantic errors. The reliability of the instrument was .96, which is considered acceptable.

Keywords
teacher evaluation, validity, reliability, undergraduate education.

Introducción

Toda evaluación efectiva y justa requiere aclarar lo siguiente: *a)* los diversos criterios de evaluación que habrán de emplearse como marco de referencia para calificar la ejecución de las funciones encomendadas al personal evaluado; *b)* las fuentes de información que permitan evidenciar el cumplimiento con los criterios de evaluación; y *c)* la utilidad de los datos para la mejora continua (Cazares, 2008).

Cada vez es más evidente que la calidad educativa pasa por la calidad de la docencia que se imparte. El diseño de políticas gubernamentales y la generación de lineamientos teóricos por parte de las comunidades académicas nacionales de reconocido prestigio no son suficientes para incrementar los niveles educativos del país; corresponde a la apropiación crítica y constructiva de los docentes promover mejoras al sistema educativo desde el aula. En este contexto, la evaluación como proceso formativo y orientador del trabajo profesional puede y debe constituirse en una herramienta indispensable para incrementar las posibilidades de lograr la excelencia educativa en cualquier institución universitaria y, por ende, en todo el sistema de educación superior.

Con base en la aplicación de diferentes instrumentos de evaluación de la práctica docente, se han efectuado investigaciones en diversos países; de ellas, es importante resaltar los siguientes resultados: *a)* el interés en la enseñanza aumenta con la edad de los profesores en menoscabo de los intereses por la investigación; *b)* los docentes modifican poco su práctica de enseñanza, en especial cuando actúan en espacios poco propicios para el cambio. Se afirma que, en general, los maestros enseñan de acuerdo con la forma en que algún profesor les enseñó y son poco propensos a modificarla; *c)* los estudiantes tienen ideas definidas y estables sobre el docente; es decir, los educandos suelen mantener en distintos momentos los mismos conceptos sobre el desempeño de sus maestros; *d)* los grupos de menos de 15 alumnos tienden a evaluar mejor a sus profesores; *e)* no hay correlación entre la calificación que recibe el docente y el porcentaje de escolares aprobados. Ello cuestiona la creencia de que el alumnado califica al maestro a partir de la calificación que éste asigna al educando (Rueda y Díaz Barriga, 2004).

Por otro lado, tenemos: *a)* los estudiantes jóvenes se interesan más por las relaciones humanas del docente que por otras de sus características; *b)* los escolares privilegian la capacidad docente del maestro sobre el conocimiento que éste posea sobre su asignatura; *c)* existe una muy débil correlación entre investigación y docencia; ello significa que no por ser un buen investigador se hace mejor profesor; *d)* aunque es aconsejable el uso combinado de evaluación por otras fuentes (autoevaluación, compañeros) y otras técnicas (observación, análisis de materiales), se sigue empleando el cuestionario como herramienta (Rueda, 2004; Rueda, 2006).

Los estudios realizados por Trent y Cohen (1973) señalan como factores principales de la docencia los siguientes: claridad de organización, interpretación y explicación; fomento de la discusión en clase y presentación de diversos puntos de vista; estimulación de intereses y motivación de los estudiantes, atención e interés por los alumnos y manifestación de entusiasmo.

Otras investigaciones han obtenido las dimensiones por medio del análisis factorial de las puntuaciones en un determinado cuestionario. Para Doyle (1975), Linn, Centra y Tucker (1975), Doyle y Whitely (1976), Ab-

bott y Perkins (1978) y Marsh (1982, 1984), algunas dimensiones son más consistentes que otras, pero, por lo común, pueden considerarse dos núcleos esenciales: *a)* competencia y destrezas docentes, que incluyen las dimensiones de estructura, organización de los contenidos, coherencia y claridad de la exposición, destrezas docentes del profesor y relevancia de los contenidos; *b)* actitud del maestro, lo que abarca actitudes del docente hacia el alumno de manera individual y hacia el grupo, estimulación del estudiante, dificultad, exigencia y justicia en la evaluación del rendimiento.

Las investigaciones de Marsh y Overall (1977, 1980) obtuvieron los siguientes factores: entusiasmo-preocupación por la tarea docente, amplitud de tratamiento de los temas presentados, organización de la tarea, interacción con los alumnos, valoración del aprendizaje por parte del estudiante, adecuación de la evaluación al desarrollo de la clase y trabajo-dificultad que ha supuesto lograr los objetivos de la materia.

Abalde, Barca, De Salvador, González y Muñoz (1995) consideran las siguientes dimensiones: relación-interacción con los educandos, metodología didáctica y evaluación. Schalock, Cowart, Staebler (1993) centran la evaluación docente en la efectividad y la productividad: conocimiento docente, destrezas docentes, competencia docente, efectividad docente y productividad docente.

Después de la revisión conceptual, se siguió la metodología de ANUIES (2006) para la evaluación, considerando sus dimensiones teóricas y complementando para el caso especial de las escuelas normales con el modelo de Marsh y Overall (1977, 1980). Las diferencias que se instauran entre las universidades y las normales obedecen a *a)* la discrepancia en el tamaño y, por tanto, a la complejidad de la funcionalidad institucional que hace, por un lado, más probable la utilización de los resultados para la mejora continua y, por otro, a la organización, en particular, en la licenciatura, haciendo posible el acceso y retroalimentación de los docentes; *b)* el proceso profesionalizante, en contraste con las universidades, implica la formación pedagógica, se trata de profesores formando profesores; *c)* los alumnos que serán los evaluadores están siendo entrenados en un proceso pedagógico; así, el modelo de evaluación debe tener validez científica.

Sin modificar los parámetros de evaluación enunciados en la bibliografía, la escuela formadora de docentes requiere medidas para que los estudiantes evalúen a sus profesores con el propósito de que la institución lleve a cabo acciones para la mejora continua. Ello implica no sólo una medida válida y confiable, sino, además, un proceso analítico flexible que genere datos para el diseño de una de estrategia de cambio.

Metodología

El estudio de validación se efectuó de dos maneras: la primera, de validación por contenido, la cual involucró la participación de 20 profesores, investigadores de educación superior, que leyeron las características del marco conceptual subyacente y evaluaron los reactivos con base en su experiencia profesional, incorporando o, en su caso, eliminando contenido de la medida. En una segunda fase, se hizo una validación por constructo que permite detectar la estructura factorial de la medida y compararla con la propuesta teórica, intentando una coincidencia entre el modelo propuesto por la estadística multivariada de independencia y el marco teórico del modelo de medida. Para realizar la validación, se necesitaron 503 estudiantes de educación superior elegidos en forma aleatoria para crear una matriz de covarianza que incluyera al menos ocho sujetos por reactivo.

Procedimiento

El diseño inicial de los reactivos se desarrolló a partir de una tabla de especificaciones y, después, validado por un total de 20 maestros, los cuales analizaron las categorías propuestas en la tabla de especificaciones, proponiendo cambios y nuevos reactivos. Una vez hechas las correcciones y con el diseño de la medida ajustada y estructurada, se aplicó a 503 alumnos de la Escuela Normal. La medida se determinó mediante grupos de 10 a 15 educandos con dos sujetos que plantearon los objetivos y propusieron las reglas de operación: objetividad, centralización, concretar responsabilidad y honestidad en sus respuestas.

El tiempo para completar la medida de práctica del docente fue de 20 minutos. Los 54 reactivos para los 503 estudiantes fueron capturados en SPSS versión 15 con el objeto de emprender los análisis factoriales para el desarrollo y la validación de constructo.

Diseño de la medida de la práctica docente

A partir del modelo de Marsh y Overall (1980) y Marsh (1982, 1984), se diseñó una medida de evaluación docente que evalúa las dimensiones enunciandas en la tabla 1 pertenecientes a la subescala planeación, evaluación y actividades didácticas del docente.

El total de los 54 reactivos podrían responderse en una escala de 7 puntos de acuerdo-desacuerdo. Los reactivos se redactaron usando el modo imperativo y con el profesor como sujeto de la oración en el nivel cognitivo para hacer juicios sobre productos de desempeño que constituían el predicado del reactivo.

Resultados

Si se considera que un instrumento de medición debe contener representados a todos los ítems del dominio de contenido de las variables a medir, la validación por contenido se refiere al grado en que la medición representa el concepto medido. Es decir, el instrumento debe contener todos

Tabla 1. Características de la medida de la práctica docente

Subescala	*Dimensión*	*Reactivos*
Planeación	Cumplimiento de obligaciones	8
	Planeación	4
Actividades docentes	Conocimientos docente	11
	Habilidades didácticas	5
	Manejo de grupo	8
Evaluación	Actividades	5
	Evaluación	13
	Total	54

los aspectos o ítems del dominio de la variable que se está midiendo y se construye de acuerdo con la teoría. En función de la teoría de ANUIES (2006) y de Marsh y Overall (1980) que sustentan este instrumento de medida, se solicitó la participación de 20 directivos y profesores en la revisión de cada uno de los reactivos con la finalidad de que se indicaran omisiones, errores sintácticos y semánticos, ortográficos y se incluyeran aspectos que el claustro considerara indispensables para la medida de evaluación del desempeño docente. A partir de los comentarios de los expertos, se efectuaron cambios sintácticos y semánticos, se incluyeron reactivos en especial de actitudes y evaluación y se eliminaron sobrantes de las dimensiones de atención, dedicación y competencia docente.

La validación por constructo ofrecerá los criterios para delimitar, en cada una de las dimensiones, los quintiles o, en su caso, deciles que se aplicarán para definir criterios de riesgo, suficiencia y excelencia dentro de la actividad pedagógica, para la planeación de la mejora continua en la actividad magisterial.

Validación por constructo

Una vez capturadas las 503 evaluaciones, se desarrolló un proceso de confianza para la captura de la base, con el resultado, en 60 instrumentos seleccionados al azar con una tabla de números aleatorios, de una captura perfecta y sin errores de las aplicaciones de la medida.

Posteriormente, se ejecutó el análisis de discriminación por reactivos, el cual consiste en sumar para cada uno de los sujetos el total de los reactivos de la evaluación y se logró un continuo con valores infinitos entre 1 y 7. Para toda la muestra, se consiguió el total de la percepción de los alumnos acerca de sus maestros. Se desagregó en cuartiles que representan a cuatro grupos donde los extremos se identifican con la percepción de profesores evaluados con excelencia (cuartil 4) y con bajo perfil (cuartil 1). A través de una *t* de student para datos independientes, se comparó para cada uno de los reactivos la media de todos los sujetos que pertenecían a los cuartiles 1 y 4.

Los datos recabados revelan que, para todos los reactivos, el cuartil 1 presentó, en todos los casos, una desviación típica mayor y una media menor que el cuartil 4. Así, en la mayoría de los casos, la prueba de Levene para igualdad de varianzas resultó significativa, por lo cual se eligieron los valores de *t* para varianzas iguales. En todos los reactivos, el valor de significancia de *t* siempre fue significativo, y se detectaron valores mínimos de 9.89 y máximos de 15.86.

En un segundo momento, se aplicó un análisis de confiabilidad con la técnica alfa de Cronbach, de donde se obtuvo un valor total de .96 para toda la prueba y valores de .91, .93, .94 y .90 para las dimensiones de didáctica, evaluación, planeación y motivación respectivamente. En la primera dimensión, las correlaciones de cada uno de los elementos con el total corregido fueron de .53 a .75, mientras que en la planeación fueron de .61 a .81; en evaluación, las puntuaciones fueron de .42 a .75. En motivación, el valor más alto de intercorrelación corresponde a .70 y el más bajo a .56.

Por último, se efectuó un análisis factorial exploratorio a través de componentes principales, ya que los dos estadísticos para valorar la bondad de ajuste al modelo factorial apoyaban tal decisión. La medida de adecuación muestral de Kaiser-Meyer-Olkin alcanzó un valor de .96 y la prueba de esfericidad de Barlett rechazó la hipótesis nula de que la matriz de correlaciones observadas es, en realidad, una matriz de identidad, asumiendo que los datos provienen de una distribución normal multivariante. El análisis de varianzas, covarianzas y antimagen indicó que las variables incluidas en el análisis comparten gran cantidad de información debido a la presencia de factores comunes, pues las correlaciones parciales son menores a .25 y, por ello, se supuso que los factores únicos son independientes entre sí; con base en ello, se decidió trabajar con rotación varimax. El gráfico de sedimentación (prueba de sedimentación de Cattell) que exhibe una representación de la magnitud de los autovalores señala que en la tendencia descendente se tiene un número de cuatro factores presentes en la solución, los cuales se derivan de 21 iteraciones de los algoritmos para encontrar la solución factorial final. Los cuatro factores explican 55.53% de la varianza total y ninguno de los componentes satura la varianza.

En la tabla 2, se observan las medidas que componen el instrumento de evaluación del desempeño docente desde la perspectiva de los estudiantes. En primer plano, la dimensión de "didáctica" con 15 reactivos, una varianza explicada de 14.31% en una escala que va de 1 a 7, incluye reactivos como "solicita a los alumnos elaborar un proyecto que incluya la interrelación/asociación/vinculación de teoría y práctica" con un peso factorial de .66, "solicita a los alumnos puntualidad con la finalidad de aprovechar el tiempo destinado de los contenidos" con un peso factorial de .65 y "presenta oportunamente el programa de la asignatura, incluyendo el cronograma, criterios de evaluación y significado de la evaluación" con un peso factorial de .29.

La dimensión de "evaluación" posee 15 reactivos con pesos factoriales que van de .74 a .35. Se logró una varianza explicada de 14.16% que incluye reactivos como "realiza evaluación coherente con los temas abordados en clase" con un peso factorial de .74, "corrige calificaciones cuando se requiere" con un peso factorial de .72 y con un peso factorial de .35 para los reactivos "da a conocer los criterios de evaluación" y "utiliza material de apoyo en clase (diapositivas, audio, video)".

La dimensión de "planeación" tiene 13 reactivos con pesos factoriales que van de .71 a .18. Se percibió una varianza explicada de 13.99% que incluye reactivos como "domina los contenidos de la asignatura que enseña" con un peso factorial de .71, "sigue un orden en la exposición de los temas" con un peso factorial de .71 y con un peso factorial .18 "distribuye el tiempo y las actividades para cumplir con todos los propósitos de la asignatura o unidad de enseñanza".

Tabla 2. Validación por constructo de un instrumento para evaluar a docentes de las escuelas normales de Sonora

Dimensión	*Peso factorial*		*% V.E.*	*α*
	Mayor	*Menor*		
Didáctica	.66	.29	14.31	.91
Evaluación	.74	.35	14.16	.93
Planeación	.71	.18	13.99	.94
Motivación	.71	.43	13.06	.81

Kaiser Mayer Olkyn. Alpha total: .96.

La dimensión de "motivación" incluye 11 reactivos con pesos factoriales que van de .71 a .43. Se aprecia una varianza explicada de 13.06% que abarca reactivos como "promueve el debate haciendo el papel de moderador y motivando a que los estudiantes participen" con un peso factorial de .72, "el profesor ofrece atención individual a los alumnos que la solicitan" con un peso factorial de .72 y con un peso factorial de .43 "ofrece apoyo académico a los alumnos que lo requieran".

Desde la perspectiva del estudiante, el modelo de medida de la práctica docente cumple con los criterios de consistencia, validez y confiabilidad definidos estadísticamente y es congruente con las características referidas en la teoría.

Discusión

En la concepción de Marsh (1984), la evaluación de los docentes contiene dos núcleos fundamentales: *a)* competencias y destrezas docentes, y *b)* actitudes. Tal como puede deducirse, los resultados apuntan que existe una dimensión didáctica, una evaluación, una planeación y una motivación y que ninguna de ellas satura la varianza explicada. Los pesos factoriales son adecuados y el Alpha de Cronbach, también. En los estudios de Marsh y Overall (1980), se mencionan algunos aspectos vinculados con la evaluación, la interacción con los alumnos, la organización con las tareas y otros. Tal como puede apreciarse, la composición factorial del instrumento para evaluar la práctica docente en las escuelas normales entraña una semejanza con las teorías existentes; en la didáctica, se hallan las competencias docentes, que incluyen la organización de las tareas, el conocimiento y los materiales, mientras que, en la planeación, se tiene la distribución de los tiempos y las estrategias de homogenización y procesos de ajuste curricular para mantener una misma velocidad de avance. En motivación, se encuentra el respeto y las características de las formas de interacción entre los alumnos y el profesor y viceversa, englobando aquellos aspectos asociados con la participación, el análisis crítico, la puntualidad y el trabajo en equipo.

Finalmente, en la evaluación, se ubican todas aquellas tareas que implican la calificación, lo mismo que el proceso de evaluación, y suponen las formas, los tipos, las asignaciones y la valoración de las diferentes tareas para lograr el desempeño deseado; así pues, aun cuando la solución factorial contenga dimensiones que parecerían lejanas de las que el modelo teórico de Marsh y Overall (1977) estipula, los contenidos son muy semejantes, sólo que, a más de diez años de distancia, los conceptos empleados en este artículo pretenden ser más generales y ajustarse al entendimiento propio de una comunidad académica que sólo se dedica a la formación de docentes para la educación básica y que, por tanto, requiere conceptualizaciones apropiadas a su contexto formativo. Por otro lado, en la validez de contenido pudo notarse el arraigo de los profesores a los modelos de evaluación docente ya existentes y que incorporan aspectos referidos en las teorías de Abbott y Perkins (1978) y en las actuales relacionadas con el diseño curricular, como las de Briggs (2005) Gagné y Briggs (2006).

En general, podríamos considerar que la valoración de contenido empata con las dimensiones halladas y el instrumento promete ser una herramienta fundamental en los procesos de mejora continua de la calidad docente en las escuelas normales.

De manera inmediata, se hace necesaria la aplicación de dicha medida al conjunto de los estudiantes de las escuelas normales para decidir los parámetros, intervalos de confianza, desviaciones estándar y otros estadígrafos que nos permitirán establecer por lo menos cinco rangos para la clasificación del desempeño docente desde la perspectiva del alumno y para cada uno de ellos colocar procedimientos remediales y, en su caso, recompensas para el comportamiento apropiado. La aplicación de una medida con tales características aunada a la autoevaluación y la evaluación de los directores podría constituir un avance siempre que se delimitaran en forma colegiada los puntos de corte en una distribución con base en la percepción de los estudiantes, desde una perspectiva basada en una medida que posibilite identificar las características académicas de los profesores, diagnosticar fortalezas y debilidades y retroalimentar al sistema la gestión de la mejora continua.

Referencias

Abalde, E.; Barca, A.; De Salvador, X.; González, R. y Muñoz, J. (1995). *Informe de avaliación da docencia universitaria polos alumnos*. Curso 93-94 y 94-95. Universidade da Coruña.

Abbott, R. y Perkins, D. (1978). Development and construct validation of a set of student rating of instruction items. *Educational and Psychological Measurement*, 38, 1069-1075.

ANUIES (2006). *Consolidación y avance de la educación superior en México. Elementos de diagnóstico y propuestas*. México: ANUIES.

Briggs, S. (2005). Changing roles and competencies of academics. *Active Learning in Higher Education*, 6(3), 256-268.

Doyle, K. (1975). *Student evaluation of instruction*. Lexington, MA: Lexington Books.

——— y Whitely, S. (1976). Implicit theories in student ratings. *American Educational Research Journal*, 13, 241-253.

Cazares, L. (2008). *Planeación y evaluación basadas en competencias: fundamentos y prácticas para el desarrollo de competencias docentes, desde preescolar hasta el posgrado*. México: Trillas.

Gagné, R. y Briggs, L. (2006). *La planificación de la enseñanza: sus principios*. México: Trillas.

Linn, R.; Centra, J. y Tucker, L. (1975). Between, within and total group factor analysis of student ratings of instruction. *Multivariate Behavioral Research*, *10*, 277-288.

Marsh, H. (1982). Validity of student's evaluations of teaching effectiveness: Cognitive and affective criteria. *Journal of Educational Psychology*, *74*, 254-279.

——— (1984). Student's evaluation of university teaching: dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology*, *76*, 707-754.

——— y Overall, J. (1977). *The Relationship between Student's Evaluations of Faculty and Instructional Improvement*. University of California. Los Angeles (ERIC Document Reproduction Service nº 138165).

——— (1980). Validity of students evaluations of teaching effectiveness: cognitive and affective criteria. *Journal of Educational Psychology*, *72*, 468-475.

Rueda, M. (2004). La evaluación de la relación educativa en la universidad. *Revista Electrónica de Investigación Educativa*, *6* (2), 1-10.

——— (2006). *La evaluación de la labor docente en el aula universitaria*. México: IISUE/ UNAM.

——— y Díaz Barriga, F. (2004). *La evaluación de la docencia en la universidad. Perspectivas desde la investigación y la intervención profesional.* México: CESU-UNAM/Plaza y Valdés, parte dos, pp. 98- 154

Schalock, M. D.; Cowart, B. y Staebler, B. (1993). Teacher productivity revised: Definition, theory, measurement and application. *Journal of Personnel Evaluation in Education*, 8 (2), 179-196.

Trent, J. y Cohen, A. (1973). A research on teaching in higher education. En R. M. Travers (ed.). *Second Handbook of Research on Teaching*. Chicago: McNally.