



Formación Universitaria

E-ISSN: 0718-5006

citrevistas@gmail.com

Centro de Información Tecnológica

Chile

Rocha, Reynaldo

ESCALA DE OPINIÓN DE LOS ESTUDIANTES SOBRE LA EFECTIVIDAD DE LA
DOCENCIA (EOEED) EN EDUCACIÓN SUPERIOR

Formación Universitaria, vol. 6, núm. 6, 2013, pp. 13-22

Centro de Información Tecnológica

La Serena, Chile

Disponible en: <http://www.redalyc.org/articulo.oa?id=373534463003>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Escala de Opinión de los Estudiantes sobre la Efectividad de la Docencia (EOED) en Educación Superior

Reynaldo Rocha

Departamento de Redes y Proyectos de Investigación Educativa, Coordinación General de Formación e Innovación Educativa, Instituto Politécnico Nacional. Av. Wilfrido Massieu s/n esq. Luis Enrique Erro, Unidad Profesional “Adolfo López Mateos”, Zacatenco, Delegación Gustavo A. Madero, C.P. 07738 Ciudad de México, D.F. – México. (e-mail: rrochac@ipn.mx, rrocha6355@gmail.com).

Recibido Jun. 12, 2013; Aceptado Jul. 18, 2013; Versión final recibida Ago. 21, 2013

Resumen

El objetivo del trabajo que se presenta fue construir un cuestionario para valorar la efectividad de la docencia incluyendo dimensiones relevantes identificadas en la literatura científica, conformando un modelo de docencia contemporáneo y coherente tanto conceptual como filosóficamente. El cuestionario sufrió sucesivas modificaciones al someterlo al juicio de 10 expertos, al de grupos de enfoque de alumnos de nivel licenciatura, a una aplicación piloto con 214 estudiantes de diferentes unidades académicas. Se aplicó finalmente sobre una muestra aleatoria, estratificada y proporcional de 99 cursos de nivel licenciatura de 8 unidades académicas del Instituto Politécnico Nacional (México), que incluyó 1627 alumnos y 99 profesores. Los resultados de consistencia interna y de validez de constructo, que se recogen en el texto, son excelentes. El producto final es un cuestionario de 17 ítems distribuidos en 4 dimensiones: relación profesor-alumno, personalidad del docente, la evaluación del estudiante realizada por el profesor, y su método de enseñanza.

Palabras clave: evaluación del docente, eficacia del docente, enseñanza superior, evaluación del estudiante, método de enseñanza.

Scale of the Student Opinion about Teacher Effectiveness (SOTES) in Higher Education

Abstract

The aim of this work was to construct a questionnaire to assess the effectiveness of teaching including relevant dimensions identified in the literature, forming a contemporary teaching model conceptually and philosophically coherent. The questionnaire was successively amended to submit to the judgment of 10 experts, of focus groups of undergraduate students, at a pilot application with 214 students from different academic units and, finally, its application on a random sample, stratified and proportional, of 99 undergraduate courses of eight academic units of the National Polytechnic Institute (Mexico), involving 1627 students and 99 teachers. The results of internal consistency and construct validity, which are reflected in the text, are excellent. The product is a 17-item questionnaire in 4 dimensions: teacher-student relationship, personality of the teacher, student assessment by the teacher and his teaching method.

Keywords: teacher evaluation, teacher effectiveness, higher education, student evaluation, teaching methods.

INTRODUCCIÓN

En los últimos años se ha destacado la importancia de la evaluación del académico para la mejora de la calidad educativa. El carácter multifuncional del profesor universitario lleva a una evaluación multifactorial que incluye la docencia, evitando privilegiarla en tanto el personal docente desempeña otro tipo de funciones, como la gestión social, la administración o la investigación (Creus et. al., 2011; De León y Huerta, 2010; Gabalán y Vásquez, 2008; Páramo, 2008; Tirado et al., 2007). En particular, desde el punto de vista de algunos organismos internacionales y de expertos en educación, la calidad de la enseñanza radica en aquellos aspectos de inciden en el aprendizaje (Carlos, 2011; Gabalán y Vásquez, 2008; Ruiz y Schumacher, 2008; Tirado et al., 2007). En este sentido, parece haber un consenso universal: un aprendizaje de calidad contribuye de manera determinante al desarrollo y progreso económico y humano (De León y Huerta, 2010; Tirado et al., 2007).

Para De León y Huerta (2010), Gabalán y Vásquez (2008) y Gorrochotegui (2005), una formación de calidad apunta no sólo a la profesionalización en una carrera universitaria, sino también a una clara formación humana y creativa que permita enfrentar con solvencia los retos de la vida. En este sentido, para Tirado et al. (2007), una faceta del aprendizaje de calidad es su contribución a la formación cívica y ciudadana de los estudiantes. Implica enseñarles a respetarse a sí mismos y a su prójimo, así como a asumir plenamente sus responsabilidades con cada miembro de la comunidad donde vive. Se requiere despertar la consideración hacia los otros, a su integridad y dignidad como personas, a sus derechos, a su prerrogativa a opinar y ser considerados, y hacia el principio de reciprocidad. Así, un aprendizaje de calidad se hace observable cuando los egresados cumplen sus compromisos y funciones, dan cuenta de ellos y promueven la transparencia.

Desde un punto de vista académico, la calidad del aprendizaje radica esencialmente en el desarrollo de ciertas competencias intelectuales. Éstas se resumen en el despliegue de un pensamiento analítico –que incluye una actitud crítica hacia los propios actos–, en la capacidad de síntesis, de combinar y relacionar ideas aprendidas en las clases formales para analizar y describir la realidad –incluso si es ajena–, reconocer patrones, principios y tomar las decisiones que permitan anticipar los retos de la vida y de la profesión. (Carlos, 2011; Ruiz y Schumacher 2008). Sin embargo, Carlos (2011) reporta diferencias entre los profesores por disciplina. “Por ejemplo, los maestros de ciencias y matemáticas le dan gran importancia al aprendizaje del dominio factual de los hechos y principios de sus disciplinas; en cambio, los de humanidades y ciencias sociales otorgan mayor peso al desarrollo personal del estudiante, a la discusión y a las habilidades comunicativas y sociales.” (p. 130). Para este último autor, todo lo anterior se trata de propósitos amplios y ambiciosos que la mayoría de los estudiantes parece no lograr. “Para remediar lo anterior se requiere un cambio profundo en las formas de enseñar” (p. 131). Donde el desarrollo continuo, actualización y compromiso del maestro es uno de los factores esenciales para mejorar la calidad de la enseñanza (Carlos, 2011; Lobato y Madinabeitia, 2011).

Por otra parte, hay un amplio acuerdo en que el logro académico de los alumnos y su satisfacción con la enseñanza recibida, son dos de los principales indicadores de calidad docente (Carlos, 2011). Páramo (2008) reseña que en la década de 1980, a partir de lo reportado en la literatura científica de su época, Marsh y sus colaboradores se dieron a la tarea de construir un instrumento para la evaluación de la enseñanza, con un constructo que reuniera las dimensiones más relevantes para la efectividad de la docencia. Las dimensiones incluidas en su Students' Evaluations of Educational Quality (SEEQ) fueron: *aprendizaje/valor, entusiasmo del instructor, interacción con el grupo, afinidad, organización/claridad, amplitud en el abordaje de los temas, exámenes/resultados, asignaciones/lecturas y trabajos/dificultad*. La dimensión *aprendizaje/valor* alude a la percepción del alumno sobre lo aprendido en el curso en cantidad y calidad en tanto valioso, innovador e intelectualmente estimulante. El *entusiasmo*, que el profesor refleja dentro y fuera de clase, es un motor de la motivación y del interés del estudiante por la asignatura. La *interacción con el grupo* se refiere a la actitud del profesor para promover la participación de los alumnos. Considera las capacidades del profesor para invitar a los estudiantes a discutir sus ideas y compartir sus conocimientos, a realizar preguntas y obtener del profesor respuestas significativas. La *afinidad* se refiere a la competencia del docente para hacer sentir cómodos y en confianza a los alumnos, hasta el punto de que se animen a buscar con regularidad su ayuda fuera y dentro de la clase. La *amplitud en el abordaje de los temas* alude a una manera particular de concebir el aprendizaje, en la que el profesor no busca transmitir una verdad única; sino que en clase contrasta las implicaciones de varias teorías. El maestro presenta su punto de vista en comparación con el de otros y discute los desarrollos más recientes en el campo de la disciplina. *Exámenes/resultados* apunta tanto a su dimensión formativa como a la rendición de cuentas. Indaga si la retroalimentación del profesor ha sido valiosa. Se investiga si lo evaluado y calificado corresponde al contenido del curso y a las temáticas enfatizadas por el docente. *Asignaciones/lecturas* como dimensión de evaluación, alude a las lecturas y trabajos extraclase exigidos a los alumnos. Se considera el valor que los estudiantes dan a estas actividades en sí mismas y para la compresión del

contenido de la asignatura. El *trabajo/dificultad* se refiere a que los alumnos perciban que les ha exigido relativamente más trabajo de lo que consideran necesario; pero con objetivos claros que les permiten saber que se espera de ellos.

Los estudios que han utilizado esta escala han reportado que sus puntuaciones son una medida válida de la efectividad del docente, más que del diseño curricular del curso. Se las ha relacionado significativamente con el aprendizaje formal de los alumnos sobre el contenido del curso; pero sujetas a la influencia de una diversidad de variables que se considera necesario controlar (Marsh y Roche, 1997; Páramo, 2008). También, han detectado algunos perfiles de docentes, por ejemplo, que puntean alto en la dimensión *organización/claridad* pero bajo en *entusiasmo del instructor* (Marsh y Roche, 1997). Para Kuzmanovic et al. (2012) esto es algo deseable pues refleja la multidimensionalidad de la docencia.

Por su parte, el instrumento *Course Experience Questionnaire (CEQ)* –*Cuestionario sobre la Experiencia del Curso*– tiene como objetivo evaluar algunas características de las asignaturas que, según la investigación educativa, promueven en los alumnos actitudes profundas a la hora de enfrentar las tareas y actividades de aprendizaje. Una de sus versiones mejor validadas, con 25 ítems, data de los 90's. Su validez y confiabilidad se comprobó empíricamente sobre una muestra nacional no probabilística del sistema de educación superior australiano y de Reino Unido. Las dimensiones de evaluación del instrumento son: *buenas enseñanzas, metas claras, evaluación apropiada, carga de trabajo apropiada y desarrollo de competencias genéricas*. Las preguntas del instrumento se miden en una escala Likert, asignando valores enteros entre 1 y 5 a posibles respuestas que van desde *definitivamente de acuerdo* hasta *definitivamente no está de acuerdo*. Originalmente fue un instrumento para la evaluación de programas de estudio completos, que se enviaba por correo a los alumnos que se hubieran graduado en el semestre inmediato anterior. Así lo reseñaron Ruiz y Schumacher (2008).

Estos autores adaptaron el CEQ para evaluar la docencia de cada profesor junto con la asignatura que impartía. Aplicaron el instrumento en una facultad de humanidades de una universidad colombiana, en seis semestres consecutivos de 2003 a 2005. Observaron que cada dimensión de evaluación del instrumento presentó un patrón semejante de fluctuación, aún comparando asignaturas de distinto tipo. Ellos interpretaron que posiblemente el instrumento registró tendencias generales de la institución y del alumnado, a pesar de preguntar específicamente por la docencia del profesor. Esta observación debilita la confianza en la validez de sus resultados. En este contexto, descubrieron una interesante relación: una *buenas enseñanzas* apoya el desarrollo de *competencias genéricas*, siempre y cuando la carga de trabajo sea percibida por los estudiantes como algo más alta de lo deseable.

En la Facultad de Estudios Superiores Iztacala de la Universidad Nacional Autónoma de México, el Consejo Técnico, integrado con representantes de profesores y alumnos de la institución, aprobó el Cuestionario de Opinión de los Alumnos sobre el Desempeño Docente en 1996. Una de sus versiones más refinadas era una escala Likert de 27 preguntas, distribuidas en cuatro dimensiones: el aprecio del alumno por la asignatura, las características generales de la docencia del profesor, características de la docencia en asignaturas de carácter teórico, características de la docencia en asignaturas de carácter práctico. Tirado et al. (2007) encontraron que los aspectos con mayor correlación con una opinión positiva del quehacer docente fueron: la claridad en la exposición, la impartición de clases interesantes (ameno) y la atención a las dudas de los alumnos. Los aspectos que fueron considerados menos relevantes fueron: el dominio del profesor de la asignatura (aunque si se estimó importante), la puntualidad y la asistencia a clases. Como medida de confiabilidad se obtuvo 0.947 de alfa de Cronbach, sobre una muestra de 25,877 alumnos ($n = 25,877$).

Según lo reportado por Gabalán y Vásquez (2008), la Universidad Autónoma de Occidente de Colombia aplicó un cuestionario de opinión de los alumnos sobre la docencia, mediante una escala tipo Likert con niveles de desempeño codificados en números enteros del 1 al 5. Los componentes evaluados fueron: responsabilidad y compromiso institucional, metodología de enseñanza, dominio de la asignatura y evaluación del desempeño de los alumnos. El instrumento fue integrado de 15 preguntas. Estos autores encontraron diferencias dependientes del tipo de asignatura en la relevancia de cada dimensión con relación a la buena docencia. A diferencia de lo reportado por Marsh y Roche (1997), no se trató de un estudio longitudinal donde se comparara la docencia del mismo profesor en distintas asignaturas. Por lo que estas diferencias se pueden deber a sesgos propios de la formación universitaria de los que suelen impartir cada tipo de curso; algunas más cercanas al ejercicio profesional de la docencia en educación superior que otras. Sin embargo, para Kuzmanovic et al. (2012) éstas pueden ser atribuibles a la heterogeneidad en las preferencias sobre los aspectos que caracterizan a la buena docencia.

Según Tirado et al. (2007), cuando se hace una revisión de los estudios de evaluación de la docencia se encuentra un alto grado de coincidencia en las dimensiones de evaluación. Entre ellas están: la sensibilidad

y preocupación del profesor por el grupo y su progreso, la preparación y organización del curso, el dominio de la asignatura, la estimulación del interés por la disciplina, el entusiasmo por la docencia, la claridad y comprensibilidad de la clase y exposiciones del profesor, las cualidades de la relación profesor-alumnos. Sin embargo, no hay coincidencia entre profesores y alumnos en la relevancia que le otorgan a cada una. En México, la mayoría de los profesores considera que lo más relevante es el dominio de la asignatura; mientras que para los alumnos es la claridad y comprensibilidad de la clase y exposiciones del profesor.

Para Kuzmanovic (2012) entre estas dimensiones se encuentran: la claridad de la clase, la interacción profesor-alumno, la organización del curso, el entusiasmo del profesor por la docencia, la retroalimentación que ofrece a los alumnos. Según los resultado de la investigación de Gabalán y Vásquez (2008), las dimensiones de evaluación que, desde la perspectiva de los alumnos, mejor discriminan un desempeño docente de calidad son: las competencias didáctico-pedagógicas del docente –que éste facilite el aprendizaje–; evaluación del aprendizaje –que las calificaciones asignadas sean justas–; y, la atención clara y precisa a las dudas y preguntas de los alumnos.

Otra dimensión de evaluación relevante, la constituyen los hábitos y actitudes del profesor. Se consideran sobre todo aquellas que se han identificado como factores que facilitan el aprendizaje. Entre los hábitos del maestro que los estudiantes aprecian de sus docentes están: “tratarlos como iguales, sonreír y tener una actitud amigable, ser accesible en horas diferentes a las destinadas para la atención en la facultad, permitir encuentros con los estudiantes fuera de las horas de clase y estar disponible antes y después de clase” (Gabalán y Vásquez, 2008, p. 16).

Para Carlos (2011), algunos principios de una enseñanza efectiva en educación superior son: despertar el interés y los deseos de aprender en los estudiantes, donde ellos acepten el esfuerzo que va requerir; preocupación y respeto por el aprendizaje del alumno, es decir, orientada a que cambie su comprensión del mundo; ofrecer una retroalimentación adecuada y una justa evaluación; clarificar frente al estudiante lo que se espera de él, y alcanzar esa finalidad debe implicar un desafío; promover la autonomía intelectual y la autorregulación del alumno; aprender de los estudiantes, lo que implica que el maestro debe ser humilde y estar dispuesto a conocer cosas nuevas; la claridad, que implica presentar el contenido de manera lógica, de lo simple a lo complejo, utilizar ejemplos y contestar adecuadamente las preguntas de los alumnos; indagar sobre las experiencias de los estudiantes y utilizarlas en la enseñanza; contrastar las implicaciones de varias teorías, incluso contrapuestas; destacar los puntos cruciales del tópico en cuestión; hacer cambios sobre la marcha de la clase en función de las actitudes de los alumnos; crear un clima favorable para su aprendizaje, de respeto y buen trato hacia ellos.

Sin embargo, debido a los estereotipos culturales, los alumnos tienen diferencias entre maestros y maestras en las dimensiones que determinan una buena docencia. Las dimensiones asociadas a las maestras mejor evaluadas fueron: apertura a las opiniones de otros, su disponibilidad para atender a los alumnos fuera de clase, la sensibilidad, su preocupación por el nivel de logro del grupo y su progreso. Mientras que las dimensiones relacionadas con los maestros hombres mejor evaluados fueron la claridad de las exposiciones y el dominio de la asignatura (Massoni y Sprague, 2005; Páramo, 2008).

Una parte de la investigación relacionada con este tipo de cuestionarios se centra en los aspectos académicos de la docencia asociados con un aprendizaje de calidad. Cuando así se confirma, se distinguen como prácticas docentes efectivas o elementos de una docencia competente. Sin embargo, varias investigaciones advierten sobre variables psicosociales –no académicas– que influyen en los resultados, algunas no atribuibles a la conducta del profesor, por lo que consideran que su validez es cuestionable (Massoni y Sprague, 2005; Páramo, 2008). Quizás por ello, cada vez más se incluyen dimensiones de evaluación relacionadas con aspectos actitudinales y de personalidad del profesor, pero abordados desde una perspectiva académica; como se ha discutido en los párrafos previos.

Por lo común, la mayoría de los instrumentos más reconocidos actualmente para recabar la opinión de los alumnos sobre el docente, se diseñaron integrando dimensiones de evaluación que se identificaron como relevantes desde la década de 1980 y anteriores. Se ha observado que algunos de estos instrumentos determinan perfiles de docencia en el sentido que dos profesores pueden obtener un puntaje equivalente pero puntuando alto en una dimensión y bajo en otra. Esto contrasta con el concepto de calidad propuesto por De la Orden et al. (2007) como el conjunto de relaciones de congruencia entre los componentes de la universidad. Extendiendo este concepto a la docencia, se traduce en la pertinencia de relaciones más congruentes entre las dimensiones que son incluidas en la evaluación que los alumnos hacen del profesor.

Desde la década de 1990, han surgido procesos que transformaron de manera significativa el quehacer cotidiano en las universidades. Entre estos se incluyen: la exigencia de acceso a la educación superior de grupos sociales que antes estaban relegados en este tipo de instituciones, la consecuente heterogeneidad

cognitiva y cultural del alumnado, una nueva economía donde las escuelas luchan por cautivar la atención de los alumnos frente a varios competidores producto de las tecnologías de la comunicación y la información (TIC's); y la crisis de financiamiento de las instituciones de educación superior.

Por eso, en este trabajo, interesó diseñar un cuestionario de evaluación del profesor sobre un modelo de docencia universitaria contemporáneo y filosóficamente consistente, como el propuesto por Gorrochotegui (2005). Esta reflexión parte de aspectos de personalidad del profesor como: la vocación por enseñar, que desde ahí se asume como un modelo y modelador de roles sociales en tanto persona que convive plenamente con otros y profesa la profesión de profesor universitario; que se ha preocupado por profesionalizar su actividad mediante el cultivo de prácticas didáctico pedagógicas intencionadas por una concepción del aprendizaje como proceso de construcción personal, compartido y negociado con otros; que implica una comprensión significativa que da lugar a cambios permanentes, conceptuales y personales que permiten interpretar la realidad en la que se está inmerso, y facilitan enfrentar con solvencia los retos y eventualidades de la vida y de la profesión.

MÉTODO Y PROCEDIMIENTO

El diseño de validación de pruebas fue basado en los trabajos de De León y Huerta (2010), García et al. (2000), Garfella et al. (2011), Marsh y Roche (1997) y McMillan y Schumacher (2005).

Población

La población del estudio está constituida por 3,302 profesores y 33,086 alumnos. Estos fueron aquello que en el segundo semestre de 2012 estaban inscritos en alguno de los 6,644 cursos, de alguno de los programas de nivel licenciatura en la modalidad escolarizada de las siguientes unidades académicas: Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME), Escuela Superior de Ingeniería y Arquitectura (ESIA), Escuela Superior de Ingeniería Química e Industrias Extractivas (ESIQIE), Escuela Superior de Ingeniería Textil (ESIT), Escuela Superior de Física y Matemáticas (ESFM) –éstas en la Unidad Zacatenco del Instituto Politécnico Nacional (IPN)–, Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas (UPIITA), ESIA y ESIME de la Unidad Ticomán del IPN. Reunidas constituyan el 60.4% de la población total en la rama Ingeniería y Ciencias Físico Matemáticas (ICFM) del instituto en la zona metropolitana de la Ciudad de México.

Muestra aleatoria proporcional estratificada

La muestra se integró con 109 de estos cursos. El primer estrato correspondió a una de las ocho unidades académicas mencionadas. El siguiente, lo constituyó los diferentes programas de nivel licenciatura en cada una. El último estrato concernió al turno dividido en *matutino* y *vespertino*. Se utilizó una lista de números aleatorios con la que se eligió la cantidad proporcional correspondiente de cursos de nivel licenciatura para cada unidad, programa académico y turno. Si algún curso correspondía al mismo grupo escolar de uno previamente elegido, fue descartado y se eligió el siguiente según la lista de números aleatorios. En la muestra proyectada, los alumnos elegidos fueron los que estuvieran inscritos en estos cursos, por lo que se tuvo cuidado de que fueran atendidos por profesores diferentes. Al levantar la muestra en campo, diez profesores rechazaron participar en esta investigación o fueron descartados por la imposibilidad de localizarlos. Además, los alumnos encuestados fueron los que asistieron a clase el día que se acordó con el profesor realizar la aplicación de los instrumentos. En total, la muestra quedó integrada por 99 profesores y 1627 alumnos. Por tipo de asignatura, la distribución de estos 99 cursos quedó como se muestra en la tabla 1.

Tabla 1: Distribución de los cursos de nivel licenciatura por tipo de asignatura ($n = 99$)

Tipo de asignatura	Porcentaje	Tipo de asignatura	Porcentaje
Matemáticas	12.1	Negocios y administración	7.1
Ciencias experimentales	23.2	Ciencias sociales y humanidades	5.1
Computación	9.1	Otras	5.0
Ingeniería	38.4		

Procedimiento

A partir de la literatura consultada, se propuso un modelo de evaluación de la docencia coherente con una concepción del aprendizaje. El aprendizaje es definido como proceso de construcción personal, compartido y negociado con otros, que implica una comprensión significativa que da lugar a cambios conceptuales y personales. Así, su calidad radica en la posibilidad de que los conocimientos adquiridos trasciendan el ámbito disciplinar, para servir al estudiante en la interpretación de la realidad en que está inmerso, adaptándose y participando en el desarrollo de una profesión y de una comunidad, cumpliendo con sus compromisos, con sus funciones y dando cuenta de ellas.

En concordancia, y con base en un trabajo previo (Rocha, 2012), las dimensiones de evaluación propuestas fueron cuatro. La primera dimensión se refirió a las características de las relaciones profesor–alumno asociadas con el fomento a la motivación del estudiante y de su nivel de involucramiento con la experiencia académica en la universidad, además de aquellas relacionadas con ubicarse frente a los alumnos como promotor –modelo– de relaciones asertivas basadas en la exigencia del respeto mutuo. La segunda fue la personalidad del profesor como ejemplo de respeto y compromiso con su profesión y con el otro. La tercera dimensión se refirió a las características de la evaluación que implementa el profesor de educación superior, incluyendo tanto el aspecto formativo como la rendición de cuentas de los aprendizajes logrados. La cuarta se refirió al aspecto didáctico pedagógico de la enseñanza universitaria. Lo que derivó en una definición operacional formada por 8 subdimensiones, 16 indicadores y 28 ítems.

En julio y agosto de 2011, la tabla de ponderación y el cuestionario fueron sometidos al análisis y evaluación de 10 jueces expertos. Los jueces evaluaron la validez de contenido de los indicadores y de los ítems, así como su inteligibilidad, mediante entrevistas a profundidad donde cada uno expresó sus opiniones al respecto. A partir de éstas, se modificó la redacción de dimensiones, subdimensiones, indicadores e ítems, se propusieron nuevos ítems y se eliminó uno, para arribar a un instrumento integrado de 37 ítems. Este instrumento se sometió a la consideración de pequeños grupos de enfoque que involucraron a 30 alumnos de diferentes unidades académicas de nivel superior del instituto. Se indagó la interpretación de los alumnos a los diferentes ítems. Considerando estas opiniones, se modificó la redacción de 16 ítems y se decidió eliminar cuatro.

Durante septiembre de 2011, se realizó una prueba piloto a esta versión integrada de 16 indicadores y 33 ítems sobre una muestra de 214 alumnos de nivel superior de seis escuelas del instituto, elegidos por el criterio de conveniencia, cercanía y acceso a los participantes. Se realizó un análisis de las cualidades métricas de los ítems. La fiabilidad interna de la escala ($\alpha = 0.966$, 32 ítems, $n = 214$ alumnos, 140 válidas); y validez de constructo mediante un análisis factorial exploratorio. Los cuatro factores extraídos concordaron parcialmente con el contenido de las cuatro dimensiones propuestas. Con base en los defectos observados, se realizaron modificaciones en la redacción de 14, se eliminaron 7 y se agregó uno.

Por otro lado, García et al. (2000) reportan que, desde la teoría clásica del test, se recomienda utilizar de 5 a 8 categorías de respuesta para una escala Likert. El dilema es el siguiente, a medida que se reduce el número de categorías se pierde información relevante para representar la variable latente; sin embargo, al aumentarlo se incrementa el número de casos conflictivos de acuerdo con el escalograma de Guttman o el modelo de respuesta graduada en el marco de la Teoría de Respuesta a los Ítems (TRI). Estos autores, mediante un proceso de modelación matemática, encontraron que el óptimo está entre 4 y 6 categorías, y es en especial recomendable 6 categorías cuando los dos factores más influyentes explican una cantidad semejante de varianza. Esto para un instrumento de 14 ítems, sobre una muestra de 1000 sujetos, con una confiabilidad de 0.85 a 0.94, y saturaciones factoriales de 0.75.

Por esto se decidió que los ítems de esta última versión del instrumento se midieran en una escala Likert asignando valores enteros entre 1 y 6 a las categorías correspondientes. Las respuestas posibles fueron: *nada de acuerdo; casi nada de acuerdo; poco de acuerdo; de acuerdo; muy de acuerdo; totalmente de acuerdo*. Se incluyó una respuesta más –*no sé o no aplica*– que se procesó como valor perdido. Este cuestionario se aplicó en las aulas, impresa y contestada a mano, del 24 de septiembre al 30 de noviembre de 2012, en una clase del profesor evaluado, obteniendo el consentimiento informado y garantizando a los participantes el anonimato de sus respuestas.

En los primeros cursos encuestados, profesor y alumnos habían convivido durante 7 semanas, que corresponde al 38.9% del tiempo curricular asignado para el total del semestre. Según el calendario escolar estos alumnos ya debían haber recibido retroalimentación del profesor derivada del primer informe de evaluación del semestre. En contraste, en los últimos cursos, profesor y alumnos habían convivido durante 17 semanas, que corresponde al 94.4% del tiempo curricular.

RESULTADOS

Se realizó un análisis factorial exploratorio por el método de componentes principales. Con base en los resultados, se decidió eliminar de la escala los 15 ítems que presentaron una comunalidad inferior a 0.600. Como resultado, en 17 ítems se encontró una estructura factorial parsimoniosa que explicó el 68.376% de la varianza; y saturaciones factoriales que oscilaron entre 0.779 y 0.882 (véase la tabla 2). Cabe notar que cada saturación factorial resultó significativa aún considerando que los datos, restando los valores perdidos, se refieren a la docencia de 98 profesores: $r(96, .01) = 0.235$.

Tabla 2: Ponderaciones de componentes para análisis factorial exploratorio ($n = 1214$ alumnos)

Dimensión	Subdimensión	Ítem	Factor	Comunalidad
Características de las relaciones profesor – alumnos.	Atención personal a los alumnos de grado.	El profesor ayuda a superar los problemas relacionados con mi aprendizaje en esta clase.	0.882	0.778
		El profesor aclara mis dudas relacionadas con el contenido de este curso.	0.846	0.715
	Gestión de los conflictos con los alumnos.	El profesor con frecuencia resuelve los conflictos que tiene con los alumnos de esta clase.	0.820	0.672
Aspectos relacionados con la personalidad del profesor como docente de grado.	Actitudes hacia el ejercicio de la docencia de grado.	El profesor de esta clase muestra vocación por enseñar.	0.860	0.740
	Ejercicio de la docencia de grado como una actividad digna y significante.	El profesor de esta clase se ha ganado mi respeto.	0.838	0.703
		Respeto al profesor como un docente comprometido.	0.834	0.695
		En esta clase, me siento escuchado por el profesor.	0.813	0.661
Características de la evaluación que implementa el profesor de grado.	Evaluación como herramienta para la mejora continua de la enseñanza y el aprendizaje.	El profesor propicia que supere los errores que cometió relacionados con el contenido de este curso.	0.844	0.712
		El profesor suele modificar la clase para mejorar mi aprendizaje del contenido de este curso.	0.819	0.671
	Evaluación para la sanción social de los resultados de los alumnos.	El profesor de esta clase asigna calificaciones que reflejan el dominio de los alumnos en el contenido del curso.	0.842	0.709
		El profesor de esta clase es justo con los alumnos al asignar calificaciones.	0.823	0.678
		El profesor de esta clase asigna calificaciones que reflejan el logro de cada alumno.	0.803	0.644
Aspectos didácticos de la enseñanza en el nivel superior.	Diseño e implementación didáctica centrados en el alumno.	El profesor toma en cuenta mis conocimientos para propiciar mi aprendizaje.	0.864	0.747
		El profesor propicia mi participación activa.	0.810	0.656
		El profesor propicia mi reflexión personal.	0.781	0.610
		El profesor conduce un repaso cuando detecta que a la mayoría de los alumnos les falta saber algo para enfrentar un tema con éxito.	0.779	0.607
	Planeación de la enseñanza.	El profesor de este curso organiza los temas en una secuencia lógica.	0.791	0.626

Coefficiente de Cronbach

Se utilizó este coeficiente para determinar la confiabilidad de la escala. Con 17 ítems, resultó: $\alpha = 0.971$ ($n = 1214$ alumnos). Mientras que las correlaciones ítem-total resultaron entre 0.751 y 0.864. De acuerdo con Marsh y Roche (1997), la confiabilidad de este tipo de escala se determina más adecuadamente a partir del consenso entre los alumnos del mismo curso. La correlación entre las respuestas de dos estudiantes de la misma clase (es decir, la confiabilidad de un solo evaluador) es típicamente entre 0.20 y 0.29, pero la confiabilidad promedio de un grupo depende del número de estudiantes: 0.95 para 50 estudiantes, 0.90 para 25 dicentes, 0.74 para 10 y 0.60 por 5 alumnos. Como se puede apreciar (véase la tabla 3), la confiabilidad de esta escala se compara favorablemente con la de las mejores pruebas objetivas.

Los 17 ítems estaban presentes en la versión original sometida a revisión piloto. Diez ítems (59%) tenían la redacción que hasta ahora se conserva. Los 7 restantes (41%) sufrieron modificaciones que mejoraron sus valores de ajuste con el modelo (estructura factorial parsimoniosa con saturaciones factoriales superiores a 0.400, sólo significativas en el factor correspondiente; con comunalidades superiores o iguales 0.600). Los

ítems eliminados fueron diez. Nueve de ellos estaban presentes en la versión piloto. Siete ítems fueron modificados en su momento porque presentaron valores de ajuste fuera de rango; acción que resultó infructuosa. Dos más habían presentado valores de ajuste adecuados en el proceso de piloteo, pero fueron descartados en esta fase de validación de la prueba. El ítem que se agregó, resultó eliminado por presentar valores de ajuste fuera de rango.

El rango teórico de esta escala de 17 ítems es de 85 puntos, desde un mínimo de 17 hasta un máximo de 102. Datos que se replicaron empíricamente. Su media fue de 76.81, 79 fue la mediana y la moda 102. Por lo que el grado de discriminación de esta escala resultó muy conveniente. Aunque, su distribución se alejó de la normal, presentando un *efecto de techo*, que según Tirado et al. (2007), es común en este tipo de medidas.

Tabla 3: Confiabilidad por grupo según número de evaluaciones válidas ($n = 98$ cursos)

Confiabilidad	Grupos con tantas evaluaciones válidas			
	Con 2 evaluaciones	De 3 a 7	De 8 a 17	De 18 a 37
Promedio	0.982	0.882	0.942	0.938
Mínima	0.968	0.295	0.613	0.806
Máxima	0.995	0.992	0.982	0.969
Total de grupos	2	22	54	20
Grupos con baja confiabilidad	0	2	1	1

DISCUSIÓN

Hasta cierto punto las modificaciones propuestas en la operacionalización del constructo fueron exitosas. Diez de los 12 ítem que presentaron valores de ajuste aceptables durante el proceso de piloteo, siguieron conservando esta condición. Siete de 14 catorce ítems cuya redacción fue modificada, mejorando sus valores de ajuste con el modelo. Además, la estructura de dimensiones y subdimensiones se sostuvo, lo que contrasta con lo ocurrido como resultado del pilotaje, donde una subdimensión se quedó sin ítems. Sin embargo, resalta la extracción de un factor único. Sorprende porque durante el proceso de piloteo se trajeron cuatro factores que concordaron con el constructo validado por el juicio de varios expertos. En efecto, en aquella ocasión se utilizó el método de rotación Oblimin, y las correlaciones entre los factores fueron entre 0.423 y 0.607, todas significativas ($p < 0.01$). Es decir, desde la versión sometida a pilotaje se presentó una gran complementariedad entre los factores, pero conservando una consistencia importante con la estructura de dimensiones propuesta desde la teoría.

En esta última versión, la estructura factorial convergió a un solo componente. Sin embargo, contrario a lo predecible para esta técnica estadística, el porcentaje de varianza aumentó de 63.715% a 68.376%. Lo que confirma que este único factor es una explicación del fenómeno que se ajusta mejor con la evidencia empírica recabada. Además, una estructura unifactorial concuerda más con una variable latente de rasgo único bien definida (unidimensional). Lo más probable es que lo obtenido en el pilotaje sea debido a un defecto de diseño que afectó su validez interna y externa (McMillan y Schumacher, 2005). En efecto, a pesar de haber incluido 214 alumnos en el pilotaje del instrumento, estos correspondieron sólo a 9 profesores. Es decir, los resultados se refirieron sólo a 9 objetos de estudio diferentes. Esto fue notoriamente insuficiente para establecer el comportamiento métrico de la escala sobre la población en estudio para este tipo de cuestionario. En cambio, en el caso del resultado final que aquí se reporta, para una escala de 17 ítems, con un rango teórico y empírico de 85 puntos, la presencia de datos referidos a la docencia de 98 profesores distintos resultó suficiente.

Por otro lado, atendiendo críticamente a las tradiciones arraigadas en este tipo de diseño metodológico, se podría argumentar que correlaciones ítem-total de 0.751 a 0.864 nos revelan redundancia entre los ítems; que quizás esté afectando la validez del constructo, es decir, que en lugar de contener 4 dimensiones conceptualmente diferenciadas se proponga por defecto una sola.

Por supuesto, hay una clara redundancia de contenido entre los ítems: “*El profesor de esta clase se ha ganado mi respeto*” y “*respeto al profesor como un docente comprometido*”; así como entre los ítems de la subdimensión “*evaluación para la sanción social de los resultados de los alumnos*”. Además, se observa que el ítem: “*En esta clase, me siento escuchado por el profesor*”, asociado con la subdimensión “*ejercicio*

de la docencia de grado como una actividad digna y significante”, bien podría vincularse con la subdimensión “atención personal a los alumnos de grado”. Esto representa la inconsistencia más clara en el contenido de la operacionalización del constructo.

Se observa que, aún eliminando las inconsistencias y redundancias, se conserva la estructura de dimensiones y subdimensiones. Dicha estructura abarca conceptos tan disímiles como aspectos casi espirituales de la personalidad del profesor relacionados con su vocación docente, pasando por asuntos actitudinales y habilidades sociales, hasta temas muy específicos de metodología de enseñanza y evaluativa. Lo que permite considerar que esta condición no revela en este caso una debilidad en la validez de constructo. Por lo tanto, a diferencia de lo que prevalece en la literatura, se ha podido construir un cuestionario para evaluar un modelo de docencia universitaria coherente, que forma parte de un todo, en lugar de sólo partes posiblemente incongruentes entre sí.

En otro asunto, al revisar la literatura, se observó que las dimensiones relevantes en la valoración de la docencia universitaria se agrupan en cuatro categorías: la experticia y dominio del profesor sobre lo que enseña; sus competencias didáctico-pedagógicas; la evaluación del desempeño de los alumnos; y aspectos de personalidad, actitudinales y socioafectivos relacionados con la posibilidad de establecer una relación que motive a los alumnos a involucrarse en la experiencia académica en la universidad. Como se observa, se ha omitido en este constructo una referencia explícita al dominio del profesor sobre la asignatura. Se consideró lo anterior debido a la cuestión pronunciada por Tirado et al. (2007) sobre qué tanto debe saber el alumno para poder valorar la experticia del docente.

Con respecto a las limitaciones de validez interna del diseño, es necesario mencionar que el periodo de aplicación de los cuestionarios es tan amplio que pudo influir en la apreciación de los alumnos sobre la docencia del profesor. Sin embargo, Marsh y Roche (1997) afirman que algunas investigaciones han obtenido correlaciones de 0.83 con calificaciones por los mismos estudiantes cuando se evaluaron de nuevo las mismas clases retrospectivamente varios años más tarde o al menos un año después de la graduación. Esto, permite prever que este defecto en el diseño pudo tener efectos razonablemente restringidos en la validez de estos resultados.

CONCLUSIONES

Se ha podido establecer una escala de 17 ítems para recabar la valoración de los alumnos sobre el docente, integrando cuatro de las dimensiones reportadas en la literatura y asociadas a su efectividad: relación profesor-alumno, personalidad del docente, la evaluación del estudiante realizada por el profesor, y su método de enseñanza. A partir de unos indicadores que han derivaron en una estructura de factor único, se permite prever que un docente bien evaluado en una de éstas lo será en las demás. Por tanto, el modelo de docencia que subyace al constructo propuesto presentó la congruencia buscada.

También, el modelo que lo sustenta está centrado en el aprendizaje, compatible con el postulado en el modelo educativo institucional según se consigna en Instituto Politécnico Nacional (2004). La medida presentó una confiabilidad comparable con la de las mejores pruebas objetivas, por lo que es recomendable su implementación en un sistema de evaluación que tenga consecuencias sobre las trayectorias docentes en la institución y la promoción escalonaria. Necesariamente, sólo como parte de uno más amplio que incluya otras miradas que recojan aspectos que se han dejado fuera, como la experticia del profesor en la disciplina.

REFERENCIAS

- Carlos Guzmán, J., *La calidad de la enseñanza en educación superior. ¿Qué es una buena enseñanza en este nivel educativo?*, Perfiles educativos: 33, pp. 129-141 (2011).
- Creus, A., Padilla Petry, P., y Sancho, J. M., *Docencia, investigación y gestión en la universidad: Una profesión tres mundos*, Praxis educativa: 16(14), pp.17-34 (2011).
- De la Orden Hoz, A., Asencio Muñoz, I., Biencinto López, C. M., González Barberá, C., y Mafokozi Ndabishibije, J., *Niveles y perfiles de funcionalidad como dimensión de calidad universitaria. Un estudio empírico en la Universidad Complutense*, Archivos Analíticos de Políticas Educativas, ISSN: 1068-2341 (en línea), 15(12), 2007. <http://epaa.asu.edu/epaa/>. Acceso: 8 de abril (2010).
- De León, A. T., y Huerta, J. M., *Cuadro de mando integral para el diseño y validación de instrumentos para valorar el desempeño académico de docentes*, Avaliação, Campinas; Sorocaba, SP: 15(1), pp. 57-72 (2010).

- Gabalán Coello, J., y Vásquez Rizo, F. E., *Del otro lado de la pizarra: relación estudiante-profesor desde perspectivas disciplinares*, Educación y Educadores: 11(1), pp.103-126 (2008).
- García Cueto, E., Hernández Baeza, A., y Muñiz, J., *Comportamiento del modelo de respuesta graduada en función del número de categorías de la escala*, Psicothema: 12(2), pp.288-291 (2000).
- Garfella Esteban, P. R., Gargallo López, B., Fernández March, A., y Suárez Rodríguez, J., *El cuestionario CEMEDEPU. Un instrumento para la evaluación de la metodología docente y evaluativa de los profesores universitarios*, Estudios sobre la Educación: 21, pp.9-40 (2011).
- Gorrochotegui, A., *Compromisos de la docencia universitaria*, Educación y Educadores: 8, pp.105-121 (2005).
- Instituto Politécnico Nacional, *Un nuevo modelo educativo para el IPN*, 1^a edición, 65-80, Dirección de Publicaciones del Instituto Politécnico Nacional, Ciudad de México, México (2004).
- Kuzmanovic, M., Martic, M., Popovic, M., & Savic, G., *A new approach to evaluation of university teaching considering heterogeneity of students' preferences*, Higher Education: doi 10.1007/s10734-012-9596-2, (2012).
- Lobato, C., y Madinabeitia, A., *Perfiles motivacionales del profesorado ante la formación en metodologías activas en la universidad*, Formación Universitaria: 4(1), pp. 37-48 (2011).
- Marsh, H. W., & Roche, L. A., *Making students' evaluations of teaching effectiveness effective. The critical issues of validity, bias, and utility*, American Psychologist: 52(11), pp. 1187-1197 (1997).
- Massoni, K., & Sprague, J., *Student evaluations and gendered expectations: what we can't count can hurt us*, Sex Roles: 53(11-12), pp. 779-793 (2005).
- McMillan, J. H., y Schumacher, S. (J. L. Posadas, Ed., J. Sánchez Baides, Trad.), *Investigación educativa. Una introducción conceptual*, 5^a edición, Pearson Educación, Madrid, España (2005).
- Páramo, P., *Factores psicosociales asociados a la evaluación del docente*, Educación y educadores: 11(1), 11-30 (2008).
- Rocha Chávez, R., *La docencia universitaria desde la perspectiva de los alumnos frente a la de los profesores*, Innovación Educativa: 12(58), pp. 91-118 (2012).
- Ruiz, F. A. y Schumacher, C., *Evaluación del aprendizaje universitario*, Educación y Educadores: 11(2), p. 91-105 (2008).
- Tirado Segura, F., Miranda Díaz, A., y Sánchez Moguel, A., *La evaluación como proceso de legitimidad: la opinión de los alumnos. Reporte de una experiencia*, Perfiles educativos: 29(118), pp. 7-24 (2007).