

Archivos Analíticos de Políticas Educativas

Revista Académica evaluada por pares
Editor: Sherman Dorn
College of Education
University of South Florida

Editores Asociados para Español y Portugués

Gustavo Fischman
Arizona State University

Pablo Gentili
Laboratorio de Políticas Públicas
Universidade do Estado do Rio de Janeiro

Volumen 14

Número 34

Diciembre 18, 2006

ISSN 1068-2341

Factores de sesgo asociados a la validez de la evaluación docente universitaria: un modelo jerárquico lineal

Raziel Acevedo Álvarez

y

Nuria Mairena Rodríguez

Universidad de Costa Rica, sede Guanacaste

Citación: Acevedo Alvarez, R. & Rodríguez, N. M. (2006). Factores de sesgo asociados a la validez de la evaluación docente universitaria: un modelo jerárquico lineal. *Archivos Analíticos de Políticas Educativas*, 14 (34). Recuperado [fecha] de <http://epaa.asu.edu/epaa/>

Resumen: El presente estudio analiza las variables del estudiante, la clase y el profesor asociadas con el sesgo en los cuestionarios aplicados a los estudiantes en la evaluación docente universitaria. En la propuesta han participado 374 estudiantes y 29 profesores de tres departamentos de la Universidad de Costa Rica (UCR), sede de Guanacaste, con una media de edad de 19.9 para los estudiantes y de 36 años los profesores. Para el análisis de los datos se utilizaron los modelos jerárquicos lineales, una metodología cuantitativa, cuyas estimaciones permitieron comprobar que de



Los lectores/as pueden copiar, mostrar, y distribuir este artículo, siempre y cuando se de crédito y atribución al autor/es y a Archivos Analíticos de Políticas Educativas, se distribuya con propósitos no-comerciales, no se altere o transforme el trabajo original. Mas detalles de la licencia de Creative Commons se encuentran en <http://creativecommons.org/licenses/by-nc-nd/2.5>. Cualquier otro uso debe ser aprobado en conjunto por el autor/es o AAPE/EPAA. AAPE/EPAA es publicada conjuntamente por el Mary Lou Fulton College of Education at Arizona State University y el College of Education at University of South Florida y, en su versión en español y portugués, por el Observatorio Latinoamericano de Políticas Educativas (OLPED) del Laboratorio de Políticas Públicas (LPP). Los artículos que aparecen en AAPE son indexados en el Directory of Open Access Journals <http://www.doaj.org> y por H.W. Wilson & Co. Contribuya con comentarios y sugerencias a Fischman@asu.edu y/o pablo@lpp-ueri.net

todos los determinantes incluidos en el estudio, solamente, cuatro de ellos (número de estudiantes en el curso, cantidad de años en la institución, tipo de departamento al que pertenece el estudiante y expectativa de nota) estaban ligeramente asociados a este tipo de evaluación. Los resultados demuestran que estas evaluaciones son válidas y se ven poco afectadas por los elementos externos a la competencia docente. Palabras claves: evaluación universitaria, modelos jerárquicos lineales, eficacia docente, modelos de regresión, validez.

Abstract: The present study analyzes the variables that are intrinsically linked with the student, professor and class environment in relation to the university educational evaluation questionnaires. The participants in the study were 374 students with an age mean of 19.9 and 29 professors with an age mean of 36 from 3 different departments at the Universidad de Costa Rica (UCR) at the city of Guanacaste. The hierarchical lineal models were used for the data analysis, a quantitative methodology which facilitates the evaluation of the determinants which affect the results of the study. However, only four of these determinants were associated with the evaluation concerned, class size, enrolment year, department type and forecasted achievement levels. The results obtained from the study demonstrate that these kinds of evaluation are valid despite the results being slightly affected by a range of factors from externalities to teacher competence.

Key words: university evaluation, hierarchical lineal models, educational competence, regression models, validity.

Los Factores de Sesgo

Antecedentes y problema

La evaluación de la docencia universitaria es un debate intenso y prolijo que se ha desarrollado, con cierta intensidad por más de cien años, alrededor del pensamiento de que la calidad del docente es un factor importante en el éxito académico de los estudiantes universitarios.

Tal planteamiento ha generado dos campos de investigaciones dedicadas a demostrar o a refutar la validez de este tipo de evaluaciones. El primero ha intentado comprobar empíricamente los factores comunes reconocidos en el “buen profesor”, tales como organización, evaluación, relaciones con los estudiantes, entusiasmo, etc. (Centra, 1983; Feldman, 1986; Marsh y Overall, 1980; Murray, Rushton y Paunonen, 1990; Ting, 2001; Cohen, 1987; Pozo, 2000; Coffey, 2001). El segundo busca demostrar la poca o nula validez de estas evaluaciones, señalando su relación con los factores contextuales de los involucrados, tales como edad, género, años en la universidad, tipo de carrera, categoría profesional del profesor, la expectativa de nota de los estudiantes, número de publicaciones, tamaño de la clase, tipo de departamento, etc. (Mateo, 1996; Fernández, 1998; Marsh, 2000, 2001).

Con el fin de medir la eficacia del docente, estas corrientes de investigación han utilizado una serie de instrumentos que les han permitido recopilar datos para sus estudios, entre éstos, la autoevaluación, la evaluación hecha por colegas, el portafolio y la percepción de los estudiantes. De todos ellos, el instrumento más estudiado respecto a las propiedades psicométricas de fiabilidad y validez, ha sido la evaluación hecha por los estudiantes.

Sobre éstos ha recaído la mayor responsabilidad de observar la calidad docente. Sin embargo, pese a la gran cantidad de literatura generada por este tema, son muchas las investigaciones que focalizan el problema de estudio en un solo nivel, ignorando con ello la existencia de otros niveles,

los cuales están interrelacionados, compartiendo la varianza con ese primer nivel. Evidentemente, lo anterior señala que estas evaluaciones son un fenómeno multinivel porque representan la percepción que tienen los estudiantes de su ambiente educacional, lo cual es el resultado de un proceso de interacción entre los estudiantes, los profesores y la clase. Ting (2001) y Marsh (2000) apoyan esta afirmación destacando que estas evaluaciones integran múltiples determinantes de los diferentes individuos, y esos aspectos, tan representativos e importantes, solo pueden ser observados desde una perspectiva anidada o multinivel.

Tal planteamiento nos presenta las siguientes interrogantes: ¿son válidos los instrumentos de evaluación de la docencia universitaria? ¿Se ven éstos afectados por factores ajenos a la competencia docente? ¿Es la evaluación docente un fenómeno multinivel?

Para dar respuesta a estas y otras interrogantes debe hacerse un estudio comprensivo sobre el tema, tomando en cuenta la propuesta de Abrami et al. (1990) referente a las “características explicativas” de la evaluación docente universitaria anidadas de la siguiente manera: las unidades de primer nivel corresponden al instrumento de medida de la competencia docente; las unidades de segundo nivel son los estudiantes, ejecutores de la evaluación; y las unidades del tercer nivel son los profesores que son los que diseñan/dirigen el programa de enseñanza y la clase misma— la clase dicta las circunstancias de intercambio entre el profesor y los estudiantes.

Determinantes de nivel I: el instrumento de evaluación

Para recopilar la información sobre la competencia del docente universitario se adaptó un cuestionario de evaluación docente facilitado por el Centro de Evaluación Académica (CEA) de la Universidad de Costa Rica (UCR).

Este centro que inició labores en 1976, se ha dedicado a la evaluación e investigación curricular con el propósito de contribuir al mejoramiento académico mediante la evaluación e investigación de las diversas dimensiones del fenómeno educativo propio de la UCR. Este centro determina la calidad de cursos y el desempeño docente mediante la aplicación de instrumentos, tales como la hoja de evaluación de desempeño docente (completada por los alumnos), la evaluación del desempeño docente (informe del jefe), la autoevaluación del desempeño docente, y la evaluación teacher's performance (versión en inglés). Por otra parte, como dependencia de la Vicerrectoría de Docencia, el CEA brinda asesoría y colaboración a la Rectoría, Consejo Universitario y a las Unidades Académicas de toda la Universidad de Costa Rica.

La hoja de evaluación de desempeño docente es un cuestionario que contiene información multidimensional sobre la opinión de los estudiantes acerca de los diferentes aspectos del diseño del curso y su evaluación del profesor, fundamentada en los siguientes seis factores identificados con la competencia docente: organización, evaluación, presentación, entusiasmo, interacción y valoración global. A continuación desglosamos lo que cada factor persigue evaluar.

Al evaluar la organización, los estudiantes deben tomar en cuenta aspectos relacionados con la estructura general del curso, la definición previa de los objetivos de trabajo dados por su profesor(a), así como también la planificación general y particular, el dominio de los contenidos temáticos, el orden de exposición de los mismos y la claridad expositiva.

El factor evaluación se refiere a la justicia e imparcialidad que un(a) profesor(a) muestra a la hora de evaluar y calificar al estudiante, la definición de los criterios de evaluación y su relación con los temas vistos en clase. Además se refiere al tipo de feedback generado por el estilo de evaluación utilizado para medir el curso.

En la presentación se califican los métodos y recursos de instrucción utilizados para acercar el conocimiento a los estudiantes; por ejemplo, el uso de multimedia, lecturas apropiadas a cada

tema de trabajo, videos, diapositivas y transparencias. También se evalúan las técnicas pedagógicas referentes a demostraciones, clase magistral y trabajo en grupo.

El factor entusiasmo se refiere a la actitud activa y dinámica de el(la) profesor(a) en la clase, si se muestra entusiasmado al impartir sus lecciones, evidenciando con ello que se siente a gusto en ella.

Con la interacción se busca evaluar las herramientas que el docente usa para estimular y provocar la participación, el diálogo, la reflexión y el debate en clase; además se busca evaluar si el(la) profesor(a) escucha las dudas y críticas de los estudiantes, manteniendo buenas relaciones con ellos.

El factor valoración global se refiere a los criterios finales donde se resume la opinión general de los estudiantes sobre el profesor y el curso.

Para seleccionar estos factores se tomó como punto de partida, además del instrumento proporcionado por el CEA, los estudios e investigaciones realizadas por Marsh (1997, 2000), Centra (1993), Feldman (1997), García Ramos (1995) y Tejedor (1990). Estos autores coinciden en que la docencia universitaria es una actividad compleja y multidimensional, sin embargo, con algunos factores comunes como la organización, la claridad, la interacción, el entusiasmo y la evaluación, entre otros. Por tanto, en nuestro estudio hemos incorporado dichos factores como criterios de evaluación de la docencia universitaria.

Los factores fueron extraídos por medio de un análisis factorial exploratorio, explicando el 61.17% de la varianza, cuyo primer factor representa el 40.56%. Luego en un análisis factorial confirmatorio, utilizando los modelos de ecuaciones estructurales, se obtuvieron altos y moderados índices de ajuste, demostrando con ello la validez del constructo propuesto (Acevedo, 2004).

Para evaluar al docente se emplea una escala aditiva global tipo Lickert, cuya puntuación para cada ítem es un valor escogido entre el 1 y el 5, siendo el número 1 el valor menor y el 5 el valor mayor. El cuestionario está compuesto por 34 ítems de este tipo, con dos ítems de control o criterios, comunes en este tipo de instrumentos. Todos los ítems se formulan en sentido positivo. La asimilación de las dimensiones es para uso formativo. En sentido sumativo, podría utilizarse la media general de todos los ítems.

La fiabilidad del cuestionario se determinó por medio del coeficiente alfa de Cronbach, obteniéndose un alfa de 0.94, valor considerado muy alto dentro de este rango. Al dividir la muestra en dos mitades, ciencias sociales obtuvo un alfa de 0.93 y ciencias básicas 0.94. Los valores obtenidos son altamente satisfactorios y destacan la consistencia interna del instrumento (Acevedo, 2004).

Al inicio del cuestionario hay una sección destinada a reunir información sobre las características del contexto de los sujetos evaluadores: años en la universidad, género, carrera, tipo de departamento, nota esperada, edad, asistencia a clase. Luego se encuentra una sección reservada para incorporar información sobre el diseño del curso, en este espacio se toma en cuenta las variables relacionadas con la cantidad de estudiantes en clase, la obligatoriedad, el grado de dificultad y cantidad de trabajo del curso.

Incluimos el primer nivel en nuestro análisis con el fin de ilustrar un posible camino a seguir para futuras investigaciones que pudiesen incluir otras características que pertenecen a este nivel, tales como mayor información sobre el mismo instrumento, variedad en el uso de instrumentos, diferentes órdenes en los ítems, variados tiempos o semestres de aplicación, entre otros. Además, al incluir el instrumento en este nivel, se vuelve posible la validación de sus diferentes factores, necesitándose un sólo método para estudiar las características psicométricas.

Particularidades de nivel II: los estudiantes

Las unidades de segundo nivel reúnen las características de los estudiantes, es decir, su género, disciplina académica, años en la universidad, su expectativa de nota, su reacción ante las condiciones del ambiente que les rodea, y su deseo por mejorar y cambiar su estructura cognitiva.

Greenwald (1997) afirma que los profesores pueden manipular su evaluación calificando con muy buenas notas a sus estudiantes. Éstos son capaces de cambiar una mala opinión de la actividad docente si el profesor les asigna una nota alta, aunque no la merezcan. Marsh (2001) y Aleamori (1999) rechazan esta afirmación y añaden que los estudiantes evalúan mejor a sus profesores que les hacen trabajar más fuerte, mostrando mayor respeto hacia sus profesores cuando perciben una mejora en el aprendizaje, aunque ello implique trabajar más duro. Por otro lado, Marsh y Roche (1997, 2000) y Feldman (1997) hacen caso omiso del efecto producido por el grado de indulgencia del profesor. Según datos empíricos, los investigadores han encontrado correlaciones menores o cercanas al 0.20 con respecto a esta variable, lo cual significa que es un efecto muy pequeño para ser tomado en cuenta. En su lugar, dichos autores recomiendan observar la variable “expectativa de nota del estudiante.”

Por otra parte, el género ha sido considerado como una fuente potencial de sesgo. En algunos estudios se observa que las estudiantes mujeres tienden a evaluar mejor a sus profesoras que a sus profesores. Los hombres, por su parte, también muestran el mismo patrón. Tatro (1995) observa que las profesoras son mejor evaluadas que los hombres. Siguiendo la misma ruta, Aleamori (1999) anota que las estudiantes mujeres son más críticas cuando sus cursos son dictados por hombres que cuando son dictados por profesoras. Por su parte, Feldman (1993) y Centra (1993) han encontrado la existencia de cierta preferencia de los estudiantes por promediar ligeramente más alto al instructor de su mismo género. Sin embargo, para Aleamori (1999) y Freeman (1994) los resultados observados en los estudios analizados son contradictorios y concluyen que el género no es un factor de sesgo en la evaluación del docente universitario.

Una particularidad considerada en el segundo nivel es la disciplina académica a la que pertenecen los estudiantes. Algunos autores afirman que los estudiantes de las ciencias blandas tienden a evaluar más alto a sus profesores que los de las ciencias duras. D'Apollonia y Abrami (1997) encuentran evaluaciones bajas para el profesor en las ciencias duras. Centra y Creech (1976) resaltan este hecho, afirmando que los profesores de humanidades reciben altas calificaciones de sus estudiantes, no así los de ciencias naturales, quienes reciben las más bajas. Consecuentemente, los profesores que trabajan en el área de humanidades resultan mejor promediados que sus colegas de las ciencias duras. Sin embargo, según Timpson y Andrew (1997), las diferencias encontradas por estos investigadores “explican menos del 1% de la varianza de las valoraciones de los estudiantes” (p. 57), lo cual es un efecto poco significativo para ser tomado en cuenta.

La edad y los años en la universidad exhiben poca evidencia empírica acerca de su efecto en la opinión de los estudiantes. Al respecto Wachtel (1998) afirma que estas variables no han sido lo suficientemente investigadas, pues son pocos los estudios relevantes como para obtener conclusiones claras. Worthington (2002) concuerda con lo anterior al señalar que el impacto de la edad ha sido sujeto de menor importancia empírica en las investigaciones sobre la evaluación de la competencia docente universitaria.

Analizando los diferentes resultados obtenidos por los investigadores mencionados podemos concluir que: a pesar que se han encontrado algunos efectos en ciertas variables como el sexo, la rama académica y otras, los resultados no son concluyentes.

Atributos de nivel III: el profesor y la clase

El tercer nivel lo forman dos componentes. El primero lo constituyen los determinantes del profesor, es decir, el rango académico, las publicaciones e investigaciones realizadas, el género, la edad, la rama académica, su nombramiento y salario. El segundo componente son las características de la clase, es decir, la cantidad de estudiantes, obligatoriedad, dificultad y cantidad de trabajo del curso.

Estos determinantes se encuentran interconectados con el ambiente de enseñanza–aprendizaje de los estudiantes y en determinadas circunstancias, pueden modular la percepción de los estudiantes respecto de su profesor. Sin embargo, aunque estas variables son muy importantes no se cuenta con suficientes estudios empíricos para profundizar en el estudio de sus efectos.

Marsh y Dunkin (1992) y Aleamori (1999) han observado que el rango académico del profesor es un indicador de su experiencia en la enseñanza y de su logro académico, por tanto, puede influir en la opinión de los estudiantes respecto a la competencia docente. En sus estudios de investigación, estos autores han encontrado cierta evidencia de un efecto positivo del rango académico que ostenta el profesor, es decir, a mayor rango y experiencia de un(a) profesor(a), mejor evaluación por parte de los estudiantes.

Otras investigaciones han tratado de demostrar que la experiencia del profesor tiene relación con su eficacia y con lo que el estudiante aprende (Feldman, 1983; Marsh, 1980, 1987; Centra y Creech, 1976). Sin embargo, los resultados encontrados no son siempre significativos o enteramente lineales, por lo que Aleamori (1974) considera que la relación entre el rango del profesor y las evaluaciones de estudiantes es ligeramente significativa, pero no hay suficiente evidencia empírica para apoyarlo.

La variable “productividad investigadora” está muy asociada con el rango académico y es uno de los indicadores de calidad docente más utilizados por las universidades para seleccionar o premiar a sus profesores. Sin embargo, los resultados obtenidos por los investigadores no son claros. Marsh (1984) no ha encontrado relación alguna entre este indicador y la evaluación de los estudiantes. Este tema puede ser un elemento potencial de inclusión en las discusiones entre las autoridades universitarias cuando se presente el momento de replantear el rol que desempeñan algunos indicadores en el ascenso de los profesores.

Otra de las variables importantes del tercer nivel es el género. Algunos estudios han encontrado cierto grado de relación entre el género del estudiante y el del profesor universitario. Otros, por el contrario, notan diferencias antagónicas entre los investigadores. En algunos casos dichos estudios reportan que los estudiantes se sienten más cómodos con profesores de su mismo género. Por ejemplo, Basow y Silberg (1987) y Summers et al. (1996) demuestran que los estudiantes prefieren a los profesores de su mismo sexo y por tanto, los evalúan mejor que a los del sexo opuesto, evidenciando la influencia del género. Por otro lado, se afirma que el género tiene poco efecto en las evaluaciones de los profesores. Feldman (1992) en su análisis sobre los estudios de laboratorio existentes, no encontró diferencias en cuanto a la relación género del profesor–género del estudiante en la calificación de la evaluación global del profesor. En una investigación posterior, Feldman (1993) replicó sus conclusiones al encontrar que la mayoría de estudios no reportaban diferencias significativas entre los géneros, un hecho que puede evidenciar la inexistencia de sesgo por género en esas evaluaciones.

Además de las características del profesor, existen otros elementos contextuales de la evaluación docente, nos referimos específicamente al lugar donde primordialmente se desarrolla la interacción entre estudiante y profesor, es decir, la clase. Siendo que la clase es el espacio donde se

desarrolla gran parte del proceso enseñanza–aprendizaje es imprescindible que ésta también forme parte del análisis del tercer nivel en nuestro modelo.

Los investigadores han encontrado que algunas de las características de la clase a menudo limitan la selección de estrategias de enseñanza y el modo de intercambio, de comunicación e interacción entre los profesores y los estudiantes. Así, se puede pensar que, en cierta manera, las evaluaciones de los estudiantes pueden reflejar parcialmente su experiencia específica en una determinada clase. Con toda seguridad, las experiencias de clase pueden estar delimitadas por la naturaleza del curso o por su obligatoriedad. Por tanto, un profesor requiere de diferentes estrategias de articulación para cada curso y debe ajustar las actividades de acuerdo con las condiciones de éstos. Claro está, en su desarrollo se requiere de las mejores habilidades y de una planificación cuidadosa para lograr hacer la materia atractiva e interesante a los estudiantes universitarios. Craton y Smith (1986); Feldman (1978, 1997); Marsh (1980, 1983, 1997); y Murray (1990) encontraron que las calificaciones más sobresalientes son otorgadas por los estudiantes de años superiores, y no por los estudiantes de recién ingreso; y generalmente en los cursos opcionales de su carrera. En otras palabras, cuando los estudiantes cuentan con la oportunidad de seleccionar los cursos de acuerdo con sus propios intereses y necesidades tienden a evaluar mejor a sus profesores; por el contrario, cuando los cursos son obligatorios, los estudiantes no son tan generosos en la evaluación de sus profesores. Con respecto a estos resultados, Aleamori (1999) dice no estar convencido, pues en sus investigaciones no se han encontrado diferencias tan relevantes como para ser tomadas en cuenta. Lo que sí parece ser evidente es que la naturaleza de una clase varía notablemente, un mismo profesor raramente muestra una evaluación consistente entre diferentes tipos de cursos, sean éstos obligatorios o electivos.

Otra variable de la clase es el grado de dificultad de la materia. Las investigaciones han tratado de comprobar la hipótesis que los profesores menos exigentes con sus estudiantes son mejor evaluados. Es decir, cuando el docente incrementa la dificultad del curso, los estudiantes pueden mostrarse insatisfechos con la clase y evaluar pobremente al profesor, penalizándolo por su grado de exigencia. Aunque esto podría parecer un razonamiento lógico para algunos, los resultados empíricos demuestran una relación totalmente opuesta. Marsh (1982, 1984) y Centra y Creech (1976) encontraron que los profesores mejor evaluados por sus estudiantes son aquellos que exigen un esfuerzo mayor para la aprobación del curso. Brady (1989) encontró que entre dos profesores altamente evaluados, los estudiantes se encuentran muy satisfechos con el más exigente de los dos. En otras palabras, los alumnos distinguen bien cuáles son los profesores que les ayudan en su transformación intelectual y cuáles no. Sin embargo, por otro lado, Arnett et al. (1989) observan que el juicio del estudiante no varía en relación con la dificultad de la materia. El razonamiento dado por estos autores es que los estudiantes entienden sobradamente que las clases son diferentes (unas son más fáciles que otras), por eso no juzgan a los profesores de acuerdo con la dificultad del curso. Ting (2001), utilizando los modelos jerárquicos lineales, expresa que el nivel de dificultad del curso no tiene un gran efecto en la evaluación del profesor por parte de los estudiantes, pero sí tiene un impacto positivo en la evaluación del esfuerzo propio del estudiante, percibiendo una mayor dedicación académica en los cursos más difíciles, y menor dedicación en los de poca dificultad. Además, este autor considera que la variable en cuestión está muy relacionada con el número de años en la universidad, la cantidad de estudiantes en clase y el tipo de curso, por lo que debe ser analizada en conjunto con estas otras y no de manera individual.

La variable cantidad de trabajo asignada por el profesor, está muy relacionada con la anterior y se ha utilizado para cuestionar o negar la validez de los cuestionarios de evaluación docente. Algunos investigadores han demostrado que a mayor cantidad de trabajo, mayor satisfacción del estudiante y por ende, mejor calificación para su profesor(a). En otras palabras, esta variable es considerada como un factor relacionado positivamente con la eficacia docente. Marsh (2001) expresa

que la cantidad de trabajo (dificultad del curso, ritmo, horas de trabajo fuera de la clase) son factores relacionados positivamente con los SET, y no negativamente, como implican las hipótesis de sesgo propuestas por algunos investigadores. Greenwald (1997) no está de acuerdo con esta afirmación y destaca que un(a) profesor(a) que asigna grandes cantidades de trabajo a sus estudiantes será penalizado por ellos en la evaluación docente, otorgándole una baja calificación. Por el contrario, si un(a) profesor(a) no asigna trabajo, o éste es escaso, los estudiantes tratarán de devolver el favor con una mejor calificación en la evaluación docente.

De todas las variables del tercer nivel, el tamaño de la clase cuenta con el mayor número de investigaciones empíricas. Diversos autores han señalado que en las clases con muchos alumnos, los profesores ven limitada su oportunidad de interacción y comunicación con ellos. Además, los estudiantes perciben una reducción de su participación activa dentro del curso debido al poco tiempo dedicado a la discusión y a la atención de consultas. Como resultado, los estudiantes pueden percibir al(la) profesor(a) como alguien que no les ayuda o que no presta atención a sus problemas individuales y ello podría afectar la evaluación del profesor.

Centra (1979); Feldman (1978) y Marsh (1987a) han encontrado la existencia de una débil relación entre el tamaño de la clase y la calidad docente. Arkerhielm (1995) y Siegfried y Kennedy (1995) aseguran que el efecto del tamaño de la clase es todavía poco claro, aunque podría influenciar tanto la percepción de los estudiantes como su satisfacción en clase. En un estudio muy interesante, Fernández (1999) demuestra que la relación existente entre el tamaño de la clase y la calidad docente no es lineal como se pensaba, sino curvilínea; por lo tanto, los estudios con resultados significativos pueden haber tomado solamente un fragmento de esa línea para sus análisis, lo cual conduciría a resultados erróneos.

Se cree que en clases con muchos estudiantes hay poca interrelación estudiante–profesor. Esta falta de interacción se ve reflejada en las dimensiones de evaluación docente relacionadas con la interacción, la comunicación y las relaciones con los estudiantes. Sin embargo, Marsh (1997) apunta que el tamaño de la clase sólo ejerce influencia en las dimensiones que naturalmente están asociadas con ella, pero de ninguna manera afecta a otras dimensiones como la organización, la presentación y la evaluación.

Sintetizando lo anterior, podemos manifestar que: los resultados obtenidos en los estudios analizados respecto de las variables contextuales asociadas al sesgo en la evaluación de la docencia universitaria no son completamente concluyentes. Sin embargo, algunas de estas investigaciones encontraron pequeños efectos relacionados con las variables contextuales, esto podría ser considerado un elemento de sesgo en estas evaluaciones. Ahora, el trabajo de nuestro estudio consiste en analizar, utilizando los modelos jerárquicos lineales, aquellas variables consideradas como sesgo potencial de la evaluación docente.

Método de Análisis

Objetivos

La profunda revisión y análisis de la literatura existente sobre la evaluación de la competencia docente ha sido la base del siguiente cuestionamiento: ¿existe un grado de asociación entre las variables contextuales y los cuestionarios de evaluación docente que pueda ser fuente potencial de sesgo? A fin de dar respuesta a esta y otras interrogantes, se plantean los siguientes dos objetivos:

Primero, identificar las variables contextuales que caracterizan a los estudiantes, profesores y la clase en la evaluación de la competencia docente universitaria, a fin de elaborar un modelo jerárquico lineal en tres niveles.

Segundo, determinar a través de los modelos jerárquicos lineales, el grado de asociación entre las variables contextuales y el cuestionario de evaluación docente, para estudiar la validez del uso de este tipo de instrumento en las evaluaciones de los docentes.

Este estudio propone un modelo jerárquico debido, en primer lugar, a la estructura de los datos. Éstos se encuentran anidados en el instrumento de evaluación, los estudiantes, el(la) profesor(a) y el curso. En segundo lugar, la naturaleza de la muestra y la variedad de relaciones entre cada nivel solamente pueden estudiarse con este tipo de metodología, la estadística tradicional es de poca ayuda en este caso.

Muestra del estudio

La muestra la forman 374 estudiantes y 29 profesores de los Departamentos de Artes, Filosofía y Letras, Ciencias Sociales y Ciencias Básicas de la Universidad de Costa Rica con sede en Guanacaste. El grupo de estudiantes tiene las siguientes características: una media de edad de 19.9 años, el 49% son mujeres y el 50% hombres, la mayoría son estudiantes de primer año (el 57%) del Departamento de Ciencias Sociales y con la expectativa de obtener una nota de ocho sobre diez en su curso. El grupo de profesores tiene las siguientes características: su edad oscila entre los 31 y los 40 años, la mayoría pertenece al Departamento de Ciencias Sociales, con el grado de instructor y con pocas publicaciones. El 86% de las clases son obligatorias, el tamaño de los grupos de clase oscila entre los 16 y 30 estudiantes, con una dificultad considerada como normal. La selección de la muestra se ha realizado por medio del procedimiento de muestreo aleatorio simple, entre todos los matriculados en los cursos de la sede de Guanacaste e interesados en participar del estudio.

Características de cada nivel

Fundamentado teóricamente el modelo jerárquico lineal en tres niveles (primer nivel el instrumento, segundo el estudiante, y tercero el profesor y su curso), condensamos en la Tabla 1 las características contextuales de cada nivel.

Tabla 1

Características contextuales de cada nivel

Nivel 1	Nivel 2	Nivel 3	
Instrumento	Estudiante	Profesor	Curso
Organización	Género	Género.	Obligatoriedad
Evaluación	Años en la	Edad	Cantidad de
Presentación	universidad	Departamento	estudiantes
Entusiasmo	Tipo de carrera	Rango	Dificultad
Interacción	Departamento	académico	Cantidad de
Valoración	Nota esperada	Investigaciones	trabajo
global	Edad	publicadas	
		Salario	

La estimación y análisis

Descripción básica del modelo.

Organizados los niveles y sus determinantes, podríamos describir cada uno de los componentes de la ecuación de regresión de la siguiente manera:

$i = 1, 2, 3$ n_{jk} son las unidades de primer nivel (los ítems del instrumento).

$j = 1, 2, 3$ indican las unidades de nivel dos (los estudiantes).

$k = 1, 2, 3$ se refieren a las unidades de nivel tres (los profesores y la clase).

Fijadas las unidades de nivel (uno, dos y tres), asumimos plenamente la distribución jerárquica de los datos y procedemos a la inclusión de una variable dependiente, en este caso el puntaje dado por los estudiantes en los 34 ítems del instrumento. La variable dependiente se escribe formalmente como: Y_{ijk}

Identificados los niveles y la variable dependiente, la ecuación básica del modelo de regresión sería la siguiente:

$$y_{ijk} = \beta_{ojk} + e_{0ijk}$$

$$\beta_{ojk} = \beta_{ook} + u_{ojk}$$

$$\beta_{ook} = \beta_{ooo} + v_{ok}$$

Donde:

Y_{ijk} es el ítem (i) contestado por el estudiante (j) respecto del profesor (k), o sea la variable dependiente.

β_{ojk} es la media de respuesta del estudiante (j) respecto del profesor (k).

β_{ook} la media general entre los profesores (k).

e_{ijk} , u_{ojk} y v_{ook} representan cada uno la varianza residual de su correspondiente nivel.

$(\sigma_e^2, \sigma_{\mu_{ojk}}^2, \sigma_{\mu_{ook}}^2)$ son los efectos al azar que determina la desviación del puntaje observado del estimado, asumiendo una distribución normal cuya media es de 0 y su varianza σ^2 .

La ecuación de tercer nivel puede integrarse en una sola, definida por:

$$Y_{ijk} = \beta_{0ijk} + v_{ok} + u_{ojk} + e_{ijk}$$

No se describe la ecuación porque representa los mismos elementos descritos en los párrafos anteriores en cada uno de sus parámetros.

Estimación del modelo nulo

El modelo nulo es el primer paso en el proceso de análisis de los datos. Este determina si existe la suficiente cantidad de varianza sin explicar, como para proceder a desarrollar modelos más complejos en donde se integren las variables contextuales de cada uno de los niveles.

Inicialmente se construye una variable con todos los valores ajustados a 1, a esta la llamaremos constante. Luego se le ubica en cada uno de los niveles jerárquicos del modelo propuesto, en nuestro caso tres niveles. Introducida la constante en el modelo, se procede con la estimación usando el programa MlwiN 1.10.

En los parámetros de estimación del modelo nulo, la parte fija, es decir, la gran media de competencia docente corresponde al valor 3.119 (de cuatro puntos), con 0.127 de error estándar, con un 95% de confianza en el intervalo. El puntaje cercano al máximo posible indica la buena opinión de los estudiantes con respecto a la labor realizada por los profesores universitarios de la Universidad de Costa Rica con sede en Guanacaste.

La parte aleatoria representa la varianza residual para cada uno de los tres niveles. Esta varianza residual nos permite determinar si los valores son significativos o no. Para ello, se divide cada uno de los resultados obtenidos entre su error estándar. El producto de esta ecuación básica ha

de ser un dígito mayor que ± 1.96 , un valor recomendado a tener en cuenta para demostrar la significatividad del parámetro ($e_{0ijk} = 0.688/0.010 = 68.8$; $u_{0jk} = 0.167/0.015 = 11.13$; $v_{0k} = 0.375/0.116 = 3.23$).

Como se observa, los valores resultantes son altamente significativos, demostrando que el modelo explicaría gran parte de la varianza entre las medias de los profesores, las medias de los alumnos y las de los ítems, indicando así, que los profesores difieren entre sí en su competencia docente, los alumnos discrepan respecto a su percepción sobre el rendimiento de su profesor y los ítems les permiten a los estudiantes discriminar las diferentes dimensiones de la competencia docente incluidas en el cuestionario de opinión. Con estos datos se comprueba la existencia de suficiente varianza en el modelo, lo que permite proceder a elaborar modelos más complejos en la búsqueda de explicaciones a nuestro problema de investigación. En caso contrario, la falta de significatividad en los datos implicaría la necesidad de plantear un modelo nuevo, o cambiar la base de datos.

Análisis de los determinantes de cada nivel

Dada la significatividad de los valores del modelo nulo, procedemos con la estimación del modelo en cada nivel. El primer nivel no ha sido estimado porque no contábamos con ninguna característica del instrumento. Seguidamente, introducimos los determinantes de segundo nivel relacionados con el género, tipo de departamento, años de universidad y el resto. Hechas las estimaciones con el programa MlwiN 1.10, la Tabla 2 presenta los resultados obtenidos.

Tabla 2

Resumen de las características de nivel dos

Características de nivel 2	β_{1jk}	v_{1jk}	Significativa $\pm >1.96$
Género	0.022 (0.075)	0.040 (0.030)	No
Años en la "U"	-0.156 (0.056)	0.043 (0.018)	Sí
Departamento	-0.162 (0.065)	0.000	Sí
Esperanza de nota	0.150 (0.039)	0.000	Sí
Edad	-0.004 (0.009)	0.000	No
Edad II	-0.033 (0.123)	0.075 (0.072)	No

Nota. Fuente: Bases de datos del instrumento aplicado en la sede de Guanacaste, UCR utilizando MlWin 10.

Como se observa, los determinantes relacionados con género, edad y edad II, no son significativos, el resultado de las operaciones (0.022/0.075, -0.004/0.009, -0.033/0.123) es un valor menor a ± 1.96 . Por tanto, podemos decir que no existe asociación alguna entre estas variables y las valoraciones hechas por los estudiantes, por lo que dichas variables deben ser eliminadas del modelo.

La Tabla 2 muestra tres determinantes que son significativos, los años de universidad, el tipo de departamento y la esperanza de nota. El primero evidencia que a mayor cantidad de años en la institución, menor es la evaluación del profesor, reduciéndose un 0.156 por cada ciclo que el estudiante permanece en la universidad. De igual forma, el tipo de departamento marca una

diferencia clara entre los estudiantes que están matriculados en las ciencias blandas y los matriculados en las ciencias duras. Los últimos ofrecen calificaciones más bajas en las evaluaciones de sus profesores (-0.162) que sus compañeros de ciencias blandas. En otras palabras, los estudiantes de ciencias duras observan en sus profesores menores habilidades que sus compañeros de ciencias blandas.

La variable esperanza de nota del estudiante está positivamente asociada con la evaluación del docente, pues a medida que sube su esperanza por obtener una mejor nota en el curso, la puntuación otorgada a su profesor aumenta en 0.150 por cada unidad de medida. Esto podría indicar que los estudiantes con la esperanza de obtener una nota alta en su curso evalúan mejor a sus profesores que aquellos sin un interés definido por obtener una calificación.

Conocida la información del nivel dos, pasamos en seguida a la estimación del nivel 3 en nuestro modelo. Como ya se mencionó anteriormente, el nivel 3 está formado por una serie de determinantes relacionados con el profesor y la clase. Ambos componentes son parte del ambiente contextual de este tipo de evaluaciones y ciertas variables pueden influir en la valoración que los estudiantes hacen de sus profesores universitarios. El detalle de las estimaciones de cada característica se puede observar en la Tabla 3.

Tabla 3

Resumen de las características de nivel tres

Características de nivel 3	β_{1k}	Significativa $\pm >1.96$
Género	0.189 (0.229)	No
Edad	0.149 (0.126)	No
Departamento	-0.314 (0.272)	No
Rango	0.034 (0.061)	No
Investigaciones	0.015 (0.074)	No
Salario	-0.033 (0.097)	No
Obligatoriedad	0.085 (0.096)	No
Cantidad de alumnos	0.169 (0.081)	Sí
Dificultad	0.001 (0.025)	No
Trabajo	0.008	No

Nota. Fuente: Bases de datos del instrumento aplicado en la sede de Guanacaste, UCR utilizando MIWin 10

Estimación conjunta de los parámetros significativos

En el siguiente modelo han sido incluidas solamente las variables significativas del segundo y tercer nivel, es decir, años de universidad, esperanza de nota, tipo de departamento y cantidad de estudiantes en clase.

Al realizar el nuevo cálculo encontramos que “cantidad de estudiantes en el curso” deja de ser significativa cuando se controlan otros aspectos de los sujetos. Por lo tanto, excluimos esta variable del modelo y realizamos el cálculo nuevamente. La Figura 1 muestra el nuevo cálculo.

$$\begin{aligned} \text{respues}_{ijk} &\sim N(XB, \Omega) \\ \text{respues}_{ijk} &= \beta_{0ijk} \text{constant} + \beta_{1k} \text{anouni}_{ijk} + 0,142(0,038) \text{espnota}_{ijk} + \\ &\quad -0,146(0,062) \text{dptocarr}_{jk} \\ \beta_{0ijk} &= 3,099(0,119) + v_{0k} + u_{0jk} + e_{0ijk} \\ \beta_{1k} &= -0,116(0,042) + v_{1k} \\ \begin{bmatrix} v_{0k} \\ v_{1k} \end{bmatrix} &\sim N(0, \Omega_v) : \Omega_v = \begin{bmatrix} 0,074(0,032) & \\ 0,030(0,013) & 0,022(0,011) \end{bmatrix} \\ u_{0jk} &\sim N(0, \Omega_u) : \Omega_u = [0,154(0,014)] \\ e_{0ijk} &\sim N(0, \Omega_e) : \Omega_e = [0,688(0,010)] \\ -2 * \log \text{likelihood(IGLS)} &= 26823,430(10562 \text{ of } 10562 \text{ cases in use}) \end{aligned}$$

Figura 1. Estimación del modelo general

Nota. Fuente: Bases de datos del instrumento aplicado en la Sede de Guanacaste, UCR utilizando MIWin 10.

La Figura 1 evidencia que la media general de este modelo difiere un poco de la obtenida con el modelo nulo, dado que $\beta_{0ijk} = 3.099$ con un error estándar de 0.119 es un valor ligeramente menor que el obtenido por el modelo nulo ($\beta_{0ijk} = 3.119, 0.127$). Estas diferencias en la media general demuestran que el punto de corte ya no es el mismo con el nuevo modelo. Ahora el valor de $\beta_{0ijk} = 3.099$ se refiere al rendimiento del docente cuando los estudiantes tienen cero años en la universidad, sin ninguna expectativa de nota y pertenecen al departamento de ciencias blandas.

En el tercer nivel, la parte aleatoria muestra valores significativos entre los profesores, demostrando la existencia de varianza residual entre éstos (0.074 sd 0.032), poniendo al descubierto que los profesores son evaluados de diferentes formas por sus estudiantes, es decir, unos reciben mayor puntaje de calificación que otros, pero no todos los profesores son vistos con el mismo grado de competencia.

En los otros efectos fijos se observan dos resultados negativos y significativos. El primero corresponde a los años del estudiante en la universidad, $\beta_{1jk} = -0.116$ con un error estándar de 0.042. Esto representa una reducción del puntaje otorgado al profesor en 0,116 por cada año de matrícula en la institución. Así, por ejemplo, si un estudiante tiene tres años en una determinada carrera, la puntuación brindada al profesor se verá reducida en 0.348. Este hallazgo podría indicar que a mayor cantidad de años en la universidad, menor satisfacción con los profesores, o mayor madurez para valorar el rendimiento del docente.

La parte aleatoria de esta variable revela que esos valores no serán iguales para todos los profesores debido a la existencia de varianza residual significativa (v_{1k} 0.022, sd 0.011), lo cual implica que para algunos profesores este valor va a ser mayor que para otros. Además, se distingue en el modelo una covarianza positiva entre el intercepto y la pendiente (0.030 sd 0.013), sugiriendo que los profesores con altos interceptos tienden a verse menos afectados por los años de universidad del estudiante.

El segundo efecto fijo negativo se ubica en el tipo de departamento al que pertenece el estudiante. En este sentido, $\beta_{2jk} -0.143$ (sd 0.062) expresa que la media general de evaluación del profesor se reduce en 0.143 cuando los profesores pertenecen al departamento de ciencias duras. En

otras palabras, los estudiantes de ciencias blandas brindan mejores puntajes de evaluación a los profesores de sus cursos que sus compañeros de ciencias duras. Los resultados encontrados en este modelo son consistentes con los datos reportados por D'Apollonia y Abrami (1997) y Marsh y Roche (2000).

Un efecto fijo positivo en este tipo de evaluación docente es la variable expectativa de nota. Los estudiantes que esperan obtener buenos resultados en su curso ofrecen una mejor evaluación de su profesor que aquellos que no tienen expectativas de notas altas. La evidencia de este resultado se encuentra en $\beta_{3,jk}$ 0.142 (sd 0.038), indicando que la variable aumenta 0.142 por cada valor de la escala (de cuatro puntos). En otras palabras, los estudiantes que esperan obtener una nota superior a ocho, aumentarían la media de evaluación de su profesor a 0.248 (2×0.142). Este resultado es consistente con las investigaciones de otros autores como Marsh y Roche (1997, 2000); Feldman (1997); y Marsh, Huo et al. (1997) quienes destacan la existencia de un efecto pequeño en la esperanza de los estudiantes por obtener altas calificaciones en sus cursos.

Conclusiones

La evaluación de la docencia universitaria es una actividad compleja y multidimensional, generadora de un gran debate relacionado en primer lugar, con los predictores o dimensiones necesarias para medir la actividad del profesor universitario y en segundo lugar, con los factores de sesgo asociados con los agentes externos al profesor, tales como el género, la edad, el tamaño del curso, los años de servicio del profesor, entre otros. Este estudio se enfoca en los factores de sesgo observando las diferentes características de los involucrados, pero ordenados jerárquicamente.

Por la cantidad de variables implicadas en la evaluación docente universitaria se ha requerido el empleo de los modelos jerárquicos lineales en tres niveles. Esta metodología permite profundizar en este tipo de fenómeno desde una perspectiva anidada. Bajo esta perspectiva se identificaron en el primer nivel los factores asociados con la competencia docente universitaria (organización, evaluación, presentación, entusiasmo, interacción, valoración global); en el segundo nivel, se observaron las variables contextuales del estudiante (edad, años de estudio, tipo de departamento, género, expectativa de nota); y en el tercero, los determinantes del profesor y la clase (género, edad, departamento, rango académico, publicaciones, salario, obligatoriedad del curso, cantidad de estudiantes, dificultad del curso y cantidad de trabajo). Todos estos componentes integraron una estructura jerárquica, conformando un amplio panorama de trabajo que permite adentrarse profundamente en la problemática de la evaluación docente universitaria. Por tanto, recomendamos utilizar los modelos jerárquicos lineales en el estudio de este fenómeno educativo, a fin de poder atender globalmente la naturaleza de este problema.

En el primer nivel no se contemplaron otras características adicionales, tales como hora del día, tipo de instrumento, año o semestre de aplicación, color, etc. Además, al haberse validado previamente los factores utilizados para evaluar al docente universitario, tampoco hubo necesidad de desarrollar este proceso. Sin embargo, recomendamos que futuras investigaciones sobre este tema tomen en cuenta características de los instrumentos de evaluación docente referentes a: variedad de los instrumentos, año o semestre de aplicación, orden de los ítems u otros elementos contextuales. No obstante, aún sabiendo que no teníamos algunos de estos factores, decidimos elaborar un modelo en tres niveles para abrir un panorama más amplio a la utilización de esta metodología de análisis.

Los determinantes de segundo nivel: género y edad del estudiante no fueron estadísticamente significativos, es decir, en este estudio no se encontraron asociados con la opinión que los estudiantes tienen sobre sus profesores universitarios.

Por otro lado, se encontró una ligera relación entre los determinantes años de universidad, tipo de departamento y esperanza de nota del estudiante y la evaluación docente. Estas variables tienden a incrementar o a disminuir la evaluación hecha por los estudiantes, por tanto, se podría pensar en la existencia de sesgo. Sin embargo, a pesar de la significatividad estadística encontrada en los puntajes de estos determinantes, los valores obtenidos con la estimación son muy pequeños como para considerarlos relevantes.

En el tercer nivel del modelo, de las nueve características de contexto incluidas en el estudio (género, edad, departamento, rango, cantidad de investigaciones, salario, obligatoriedad del curso, dificultad del curso y cantidad de trabajo del curso), ocho fueron excluidas del análisis por no lograr la significatividad estadística. Un hecho que demuestra nuevamente la robusta estructura de los cuestionarios de evaluación docente. Únicamente se encontró asociación positiva con la variable número de estudiantes en el curso. Es decir, conforme crece el número de estudiantes en el curso, la evaluación positiva tiende a incrementarse ligeramente. Este hallazgo contradice los resultados de otras investigaciones donde se ha encontrado una relación negativa entre las evaluaciones de los profesores y el tamaño del curso. Una posible explicación de nuestro resultado puede ser el hecho de que en la sede de Guanacaste, Costa Rica no existen grupos mayores de 50 estudiantes por clase y además, la mayoría de los grupos con más de 31 estudiantes eran de primer año, atendidos por profesores con muchos años de experiencia y con un alto rango universitario. Ciertamente, los estudiantes se sienten muy conformes con estos profesores.

Finalmente, ante las interrogantes del estudio, se ha demostrado que el cuestionario de evaluación docente universitaria utilizado en esta investigación reúne valores altamente significativos como para considerarlo válido y confiable. Además, debido a la consistencia de su estructura, éste no es afectado por los factores de sesgo ajenos a la competencia docente, lo cual demuestra una vez más la validez de este tipo de evaluación.

En este sentido, el modelo jerárquico lineal presentado reúne condiciones suficientes para la obtención de resultados válidos en cualquier ámbito de la docencia universitaria, ya que permite la interrelación de diferentes variables contextuales, confirmando que los cuestionarios de evaluación de la docencia universitaria son un fenómeno multinivel.

Ciertamente, hemos observado que muchas de las características utilizadas como indicadores de calidad docente universitaria no tienen relación con la opinión de los estudiantes respecto de sus profesores. Para el grupo de estudiantes de nuestra muestra, la calidad docente tiene poco que ver con el número de publicaciones, el género, los años de experiencia y otros elementos contextuales. Por otro lado, sí se ha encontrado una estrecha relación entre la calidad docente y la organización de la materia, el tipo de evaluación aplicada en el curso y la interacción del profesor con sus estudiantes. Por esta razón, instamos a los administradores universitarios a cuestionar estos indicadores y a replantear otros nuevos.

A futuro, es necesario eliminar el modelo en tres niveles y experimentar con dos niveles, pero controlando cada uno de los factores que componen el instrumento de evaluación docente. Una vez controlados los factores, es necesario contrastarlos con las variables significativas de este estudio: tipo de departamento, años en la universidad, expectativa de nota y cantidad de estudiantes en el curso. Una vez elaborado el nuevo modelo, se podría observar si los pequeños efectos encontrados en esta investigación se mantienen o desaparecen cuando se controlan los factores de competencia docente universitaria.

Referencias

- Abrami, P. C., D'apollonia, S. & Cohen, P. (1990). Validity of Student Ratings of Instruction: What We Know and What We Do Not. *Journal of Educational Psychology*, 82 (2), pp. 219-231.
- Acevedo, R. & Fernández, M. J. (2004). La percepción de los estudiantes universitarios en la medida de la competencia docente: validación de una escala. *Revista de Educación* 28 (2), pp. 145-166. San José: Editorial de la Universidad de Costa Rica.
- Akerheilm, K. (1995). Does Class Size Matter? *Economics of Education Review*, 14 (3), pp. 229-241.
- Aleamori, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, 13 (2), pp. 153-166.
- Aleamori, L. M. & Graham, M. H. (1974). The relationship between CEO ratings and instructor's rank, class size and course level. *Journal of Educational Measurement*, 11, pp. 189-202.
- Basow, S. A. & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology*, 79, pp. 308-314.
- Brady, P. J. (1989). *Do students evaluate professors from a consumer-product standpoint?* Paper presented at the American Educational Research Association (AERA) Conference. San Francisco, CA.
- Centra, J. A y Creech, F. R. (1976). The relations between student teacher and course characteristic and student rating of teacher effectiveness. *Project Report 76, 1*. Princeton, NJ.: Educational Testing Service.
- Centra, J. A. (1979). *Determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Coffey, M. & Gibbs, G. (2001). The evaluation of the student evaluation of educational quality questionnaire (SEEG) in U.K. higher education. *Assessment & Evaluation in Higher Education*, 26 (1), pp. 89-93.
- Cohen, P. A. (1987). *A critical analysis and reanalysis of the multisection validity meta-analysis*. Paper presented at the American Educational Research Association (AERA) Conference. Washington, D.C. (Eric Document Reproduction Service No ED 283876)
- Craton, P. A. & Smith, R. A. (1986). A new look at the effect of course characteristics on student ratings of instruction. *American Educational Research Journal*, 23 (1), pp. 117-128.
- D'Apollonia, S. & Abrami, P. C. (1997). Navigating student ratings of instruction. *American Psychologist*, 51 (11), pp. 1198-1208.
- Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't know. *Research in Higher Education*, 9 (2), pp. 199-242.
- Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education*, 18 (3)-124.
- Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review synthesis. *Research in Higher Education*, 24, pp. 129-213.
- Feldman, K. A. (1992). College students' view of male and female college teachers: Part I. Evidence from students' social laboratory and experiments. *Research in Higher Education*, 33, pp. 317-375.
- Feldman, K. A. (1993). College students' view of male and female college teachers: Part II. Evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, 34 (2), pp. 151-191.

- Feldman, K. A. (1997). Identifying exemplary teachers and teaching: evidence from student ratings. En R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice*. Bronx, NY: Agathon.
- Fernández, J. (1999) Evaluación de la docencia y aprendizaje profesional: análisis de una experiencia universitaria. *Revista Interuniversitaria de Formación del Profesorado*, 34, pp. 87-98.
- Fernández, J., Mateo, M. & Muñoz, J. (1998). Is there a relationship between class size and student ratings of teacher quality? *Educational and Psychological Measurement*, 58, pp. 596-604
- Freeman, H. R. (1994). Student evaluation of college instructors: Effects of type of course taught, instructor gender and gender role, and student gender. *Journal of Educational Psychology*, 86 (4), pp. 627-630.
- García Ramos, J. M. (1995). La evaluación institucional a través del departamento de recursos humanos en la institución universitaria privada. *Revista Bordon*. Madrid.
- Greenwald, A. G. & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 51 (11), pp. 1209-1217.
- Marsh, H. & Roche, L. R. (1997). Making students' evaluation of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 2 (11), pp. 1187-1197.
- Marsh, H. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students evaluation of university teaching. *British Journal of Psychology*, 52, pp. 77-95.
- Marsh, H. (1983). Students's evaluation of university teaching: research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3), pp. 253-288.
- Marsh, H. (1984). Student's evaluation of university teaching; dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology*, 76 (5), pp. 707-754.
- Marsh, H. (1987a). Students' evaluation of university teaching; research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, pp. 253-288.
- Marsh, H. (1987b). Student evaluations of teaching. En M. J. Dunkins (Ed.), *The International Encyclopedia of Teaching and Teacher Evaluation*. Pp.181-187. Oxford: Pergamon Press.
- Marsh, H. W. & Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. En J. C. Smart (Ed.), *Higher education: handbook of theory and research*, Vol. 8, pp. 143-233. New York: Agathon Press.
- Marsh, H. W. & Overall, J. U. (1980). Validity of students' evaluations of teaching effectiveness: cognitive and effective criteria. *Journal of Educational Psychology*, 72 (4), pp. 468-475.
- Marsh, H. W. (1980). The influence of student, course, and instructor characteristics on evaluations of university teaching. *American Educational Research Journal*, 17, pp. 219-237.
- Marsh, H. W. (2001). Distinguishing between good (useful) and bad workloads on students' evaluations of teaching. *American Educational Research*, 38 (1), pp. 183-212.
- Marsh, H. W. & Roche, L. A.(2000). Effects of grading leniency and low workload on students' evaluation of teaching: Popular myth, bias, validity or innocent bystanders? *Journal of Educational Psychology*, 92 (1), pp. 202-228.
- Marsh, H. W., Hau, K. T., Chung, C. M. & Siu, T. L. P. (1997). Students' s evaluations of university teaching: Chinese version of the student's evaluations of educational quality instrument. *Journal of Educational Psychology*, 89 (3), pp. 568-572.
- Mateo, M. A. & Fernández, J. (1996). Incidence of class size on the evaluation of university teaching quality. *Educational and Psychological Measurement*, 56, pp. 771-778.
- Murray, H. G., Rushton, P. & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*, 82 (2), pp. 250-261.

- Pozo Muñóz, C., Roboloso, E. & Fernández, B. (2000). The "ideal teacher". Implications for student evaluation of teacher effectiveness. *Assessment & Evaluation in Higher Education*, 25(3), pp.253-263.
- Siegfried, J. J. & Kennedy, P. E. (1995). Does pedagogy vary with class size in introductory economics? *AERA Papers and Proceedings*, 85 (2), 347-351.
- Summers, M. A., Anderson, J. L., Hines, A. R., Gelder, B. C. & Dean, R. S. (1996). The camera adds more than pounds: gender differences in course satisfaction for campus and distance learning student. *Journal of Research and Development in Education*, 29 (4), pp. 212-219.
- Tatro, C. N. (1995). Gender effect on student evaluations of faculty. *Journal of Research and Development in Education*, 28 (3), pp. 169-173.
- Tejedor, F. & Montero, L. (1990). Indicadores de la calidad docente para la evaluación del profesor universitario. *Revista Española de Pedagogía*, año XLVIII, 186, pp. 259-279.
- Timpson, W. & Andrew, D. (1997). Rethinking student evaluations and the improvement of teaching: Instruments for change at the university of queensland. *Studies in Higher Education*, 22 (1), pp. 55-66.
- Ting, K. F. (2001). A multilevel perspective on student ratings of instruction: Lessons from the chinese experience. *Research in Higher Education*, 41(5), pp. 637-653.
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23, pp. 191-211.
- Worthington, A. C. (2002). The impact of student perception and characteristics on teaching evaluation: a case study in finance education. *Assessment & Evaluation in Higher Education*, 27 (1), pp. 49-64.

Datos biográficos sobre los autores

Raziel Acevedo Álvarez es doctor en Métodos de Investigación, Diagnóstico y Evaluación Educativa. Profesor asociado de la Universidad de Costa Rica, sede de Guanacaste.

Nuria Mairena Rodríguez es magistra en Administración Educativa. Profesora instructora Universidad de Costa Rica, sede de Guanacaste.

Dirección de los autores

La correspondencia relacionada con este artículo debe enviarse a:

Raziel Acevedo Álvarez. Universidad de Costa Rica, sede de Guanacaste, Liberia. Tel: (506) 680-0461, fax: (506)- 666-1206, correo electrónico: racevedo@cariari.ucr.ac.cr
Nuria Mairena Rodríguez, Universidad de Costa Rica, Sede de Guanacaste, Liberia. Tel: (506) 666-0357, fax (506)666-12-06, correo electrónico: nuria_mairena@costarricense.cr

ANEXO N° 1

Para responder a las siguientes preguntas se utilizará una escala 1 a 5, en la cual 1 es el menor puntaje y 5 el mayor. Usted puede marcar en uno de estos extremos, o en cualquiera de los puntajes que se encuentren entre ellos, dependiendo de la valoración que desea otorgar al profesor o profesora de este curso. Puede rellenar el círculo o marca con una equis (X) sobre el número que corresponda.

SU PROFESOR O PROFESORA

Evidencia que ha preparado las clases con anticipación	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Brinda la oportunidad de exponer dudas y preguntas, respondiéndolas con claridad.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Provoca el diálogo, la reflexión y debate sobre los temas tratados en clase.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Los objetivos de curso se definen anticipadamente.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Acepta críticas a sus objetivos de clase, planteamientos y otros, por parte de los y las estudiantes.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Se interesa por los y las estudiantes que demuestran problemas en aprender, los temas de estudio.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Contrasta varias implicaciones en los temas que explica.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Brinda un contexto global a sus ideas y conceptos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Abarca los temas de estudio en su totalidad.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Emplea variedad de ayudas audiovisuales (multimedia, diapositivas, videos, etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Se encuentra disponible en el horario que estableció para horas de consulta extra clase	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Muestra dominio de los temas tratados	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Utiliza variedad de recursos de enseñanza (demostraciones, lecturas, trabajo en grupo, etc.).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Es ordenado al exponer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Trata de acercarse y conversar con los estudiantes antes y después de clase.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Explica con claridad	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mantiene buenas relaciones con los y las estudiantes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Resume y enfatiza los aspectos claves de cada lección.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
El o los procedimiento (s) de evaluación permite al alumno reflejar sus conocimientos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Refleja entusiasmo en la presentación de la clase o de los materiales.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Tiene una forma de enseñar que facilita la comprensión de la materia	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Elabora los exámenes para sintetizar efectivamente la materia del curso	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mantiene mi atención en la clase	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Es justo e imparcial a la hora de evaluar y calificar (exámenes, quices, tareas, asignaciones, otros)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Se mantiene activo y dinámico cuando dirige este curso	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Devuelve las evaluaciones ya calificadas dentro de los diez días hábiles siguientes a su realización o entrega	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Asiste regularmente a clases y si falta lo justifica.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Genera una retroalimentación en los exámenes, calificaciones y materiales, que me ayudan a comprender la materia de clase.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Como consecuencia de este curso tú interés por la materia se ha incrementado	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Globalmente usted evaluaría a su profesor con	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

AAPE Comité Editorial

Editores:

Gustavo E. Fischman, Arizona State University

Pablo Gentili, Universidade do Estado do Rio de Janeiro

Asistentes editoriales: Rafael O. Serrano (ASU–UCA) & Lucia Terra (UBC)

Hugo Aboites

UAM-Xochimilco, México

Claudio Almonacid Avila

UMCE, Chile

Alejandra Birgin

FLACSO-UBA, Argentina

Mariano Fernández Enguita

Universidad de Salamanca. España

Roberto Leher

UFRJ, Brasil

Pia Lindquist Wong

CSUS, USA

Alma Maldonado

University of Arizona, USA

Imanol Ordorika

IIE-UNAM, México

Miguel A. Pereyra

Universidad de Granada, España

Romualdo Portella de Oliveira

Universidade de São Paulo, Brasil

José Ignacio Rivas Flores

Universidad de Málaga, España

José Gimeno Sacristán

Universidad de Valencia, España

Susan Street

CIESAS Occidente, México

Daniel Suárez

LPP-UBA, Argentina

Jurjo Torres Santomé

Universidad de la Coruña, España

Armando Alcántara Santuario

CESU, México

Dalila Andrade de Oliveira

UFMG, Brasil

Sigfredo Chiroque

IPP, Perú

Gaudêncio Frigotto

UERJ, Brasil

Nilma Lino Gomes

UFMG, Brasil

María Loreto Egaña

PIIE, Chile

José Felipe Martínez Fernández

UCLA, USA

Vanilda Paiva

UERJ, Brasil

Mónica Pini

UNSAM, Argentina

Paula Razquin

UNESCO, Francia

Diana Rhoten

SSRC, USA

Daniel Schugurensky

UT-OISE Canadá

Nelly P. Stromquist

USC, USA

Antonio Teodoro

Universidade Lusófona, Lisboa

Lílian do Valle

UERJ, Brasil

EPAA Editorial Board

Editor: Sherman Dorn University of South Florida

Production Assistant: Chris Murrell, Arizona State University

[Michael W. Apple](#)

University of Wisconsin

[Greg Camilli](#)

Rutgers University

[Mark E. Fetler](#)

California Commission on Teacher
Credentialing

[Richard Garlikov](#)

Birmingham, Alabama

[Thomas F. Green](#)

Syracuse University

[Craig B. Howley](#)

Appalachia Educational Laboratory

[Patricia Fey Jarvis](#)

Seattle, Washington

[Benjamin Levin](#)

University of Manitoba

[Les McLean](#)

University of Toronto

[Michele Moses](#)

University of Colorado

[Anthony G. Rud Jr.](#)

Purdue University

[Michael Scriven](#)

University of Auckland

[Robert E. Stake](#)

University of Illinois—UC

[Terrence G. Wiley](#)

Arizona State University

[David C. Berliner](#)

Arizona State University

[Linda Darling-Hammond](#)

Stanford University

[Gustavo E. Fischman](#)

Arizona State University

[Gene V. Glass](#)

Arizona State University

[Aimee Howley](#)

Ohio University

[William Hunter](#)

University of Ontario Institute of
Technology

[Daniel Kallós](#)

Umeå University

[Thomas Mauhs-Pugh](#)

Green Mountain College

[Heinrich Mintrop](#)

University of California, Los Angeles

[Gary Orfield](#)

Harvard University

[Jay Paredes Scribner](#)

University of Missouri

[Lorrie A. Shepard](#)

University of Colorado, Boulder

[Kevin Welner](#)

University of Colorado, Boulder

[John Willinsky](#)

University of British Columbia

***EPAA* English-language Graduate-Student Editorial Board**

Noga Admon

New York University

Jessica Allen

University of Colorado

Cheryl Aman

University of British Columbia

Anne Black

University of Connecticut

Marisa Cannata

Michigan State University

Chad d'Entremont

Teachers College Columbia University

Carol Da Silva

Harvard University

Tara Donahue

Michigan State University

Camille Farrington

University of Illinois Chicago

Chris Frey

Indiana University

Amy Garrett Dikkers

University of Minnesota

Misty Ginicola

Yale University

Jake Gross

Indiana University

Hee Kyung Hong

Loyola University Chicago

Jennifer Lloyd

University of British Columbia

Heather Lord

Yale University

Shereeza Mohammed

Florida Atlantic University

Ben Superfine

University of Michigan

John Weathers

University of Pennsylvania

Kyo Yamashiro

University of California Los Angeles