

Revista Mexicana de Investigación Educativa

ISSN: 1405-6666

revista@comie.org.mx

Consejo Mexicano de Investigación Educativa,  
A.C.

México

García Garduño, José María

¿Qué factores extraclase o sesgos afectan la evaluación docente en la educación superior?

Revista Mexicana de Investigación Educativa, vol. 5, núm. 10, julio-dici, 2000

Consejo Mexicano de Investigación Educativa, A.C.

Distrito Federal, México

Available in: <http://www.redalyc.org/articulo.oa?id=14001006>

- ▶ How to cite
- ▶ Complete issue
- ▶ More information about this article
- ▶ Journal's homepage in redalyc.org

## INVESTIGACIÓN

Revista Mexicana de Investigación Educativa  
julio-diciembre 2000, vol. 5, núm. 10  
pp. 303-325

### ¿Qué factores extraclasses o sesgos que afectan la evaluación docente en la educación superior?

José María García Garduño\*

#### **Resumen:**

A pesar de que el empleo de los cuestionarios de evaluación de la docencia se está tornando una práctica común en las más importante instituciones universitarias de México, poco se ha escrito e investigado en español sobre los factores extraclasses que afectan la evaluación docente en la educación superior. El objetivo de este trabajo fue realizar una síntesis de la investigación acerca de los factores extraclasses que afectan la evaluación docente en la educación superior, de acuerdo con los hallazgos de la investigación generados por el empleo de los cuestionarios de evaluación de la docencia. Se concluye que aunque los factores extraclasses o sesgos influyen en la evaluación de la efectividad docente por parte de los alumnos, los cuestionarios de evaluación son instrumentos válidos y confiables para medir la efectividad docente. Sin embargo, se sugiere cautela en el uso de los cuestionarios de evaluación docente con fines sumativos. El autor sugiere algunas recomendaciones prácticas, derivadas del estudio, para el diseño y empleo de los cuestionarios de evaluación.

#### **Abstract:**

In spite of the widespread use of student ratings of instruction in Iberoamerican higher institutions, little has published in Spanish language on the biases or background factors that affect teaching effectiveness. There are some misconceptions and, to some extent, natural resistance by professors and administrators as well on the use of student ratings of instruction to assess teaching effectiveness. The purpose of this paper was to carry out a comprehensive research review in order to ponder the actual effects of those biases or background factors on the assessment of teaching effectiveness by students. The author concludes that in spite of student ratings of instruction are affected by some biases, they are a valid and reliable source for assessing teaching effectiveness. However, the article suggests that the use of student ratings of instruction for promotion should be made with caution. The author suggests some practical recommendations to improve the use of student ratings of instruction.

**Palabras clave:** cuestionarios de evaluación de la docencia, sesgos, desempeño docente, profesores, alumnos.

---

\* Profesor-investigador del Departamento de Educación y Desarrollo Humano de la Universidad Iberoamericana, CE: josemaria@uia.mx

**Key words:** student ratings of instruction, bias, teaching effectiveness, instructors, students.

La investigación de la efectividad docente a través de los cuestionarios de evaluación ha sido uno de los tópicos más prolíficos y antiguos dentro de la educación superior. Así lo demuestran los más de dos mil artículos que se han publicado a partir de que Remmers creara el primer cuestionario de evaluación, e investigara sobre su uso en la Universidad de Purdue, a finales de la década de los veinte (Feldman, 1990; Marsh, 1987; Centra, 1993). A pesar de que el tópico ha dado mucho que investigar y que en México y en España este tipo de cuestionarios se emplearon, por primera vez, hace casi treinta años y que cada vez su uso se está volviendo más popular, tanto en universidades públicas como privadas, los trabajos que sintetizan los avances de la investigación son mínimos o ausentes en el idioma español.

Entre los profesores y funcionarios universitarios existe cierta resistencia y un conocimiento intuitivo que los lleva a descalificar la utilidad y validez de la evaluación de la docencia por medio de los alumnos. Se ha dicho que los estudiantes no son capaces de calificar la complejidad de la tarea docente, que las evaluaciones son meros concursos de popularidad y que los buenos resultados están ligados a las notas que el profesor otorga a sus alumnos. Sin embargo, existe una gran cantidad de estudios que indican que los cuestionarios de evaluación de la docencia son instrumentos confiables y válidos para valorar la efectividad docente, y que esas preocupaciones o resistencias naturales de los profesores y funcionarios universitarios para emplearlos no están del todo sustentadas en la investigación empírica. No obstante, la extensa investigación sobre el tema indica que existen variables que influyen en los resultados de la evaluación de los profesores por los alumnos. De ahí que el propósito de este trabajo sea examinar los resultados de la investigación sobre los factores extraclase que afectan la evaluación de la efectividad docente del profesor.

***¿Los factores o sesgos que afectan la evaluación del profesor por los alumnos son inherentes a la tarea docente?***

Los factores extraclase o sesgos en los cuestionarios de evaluación de la docencia por los alumnos (CEDA) se refieren a todas aquellas variables no relacionadas con la instrucción que afectan la efectividad docente en el salón de clase (Marsh, 1984). Se ha reconocido que los alumnos no evalúan al profesor con total independencia; su idiosincrasia y antecedentes, así como las características del maestro y del curso juegan, entre otros factores, un papel importante en los resultados de la valoración; a estos factores, extra-docentes que afectan la evaluación se les ha denominado sesgos, variables extrañas o antecedentes de los alumnos (Feldman, 1977; Marsh, 1984). Tal vez el sesgo o factor extraclase más reconocido por los propios profesores sea la nota final del curso; se cree que quienes otorgan las notas o calificaciones más altas reciben mejores evaluaciones por parte de los estudiantes.

La bibliografía sobre los factores extraclase o sesgos en los CEDA es abundante. Sin embargo, Marsh y colaboradores (Marsh 1987; Marsh y Roche, 1997) consideran que la mayoría de esos estudios son ateóricos y padecen de limitaciones metodológicas; los autores mencionan las siguientes:

- a) confusión de correlación con causalidad;
- b) empleo de una inapropiada unidad de análisis, el promedio del grupo es la unidad apropiada, pero no la evaluación individual del alumno;
- c) descuido en la consideración de la naturaleza multivariada de los CEDA;
- d) definición inapropiada de lo que se considera como sesgos y
- e) inapropiada manipulación experimental en los estudios de laboratorio.

A pesar de las limitaciones que señala Marsh, la bibliografía sobre los sesgos ha demostrado que sí existen factores ajenos al proceso de instrucción que afectan positiva o negativamente la efectividad del profesor. Sin embargo, podría afirmarse que, en sentido estricto, no son sesgos sino más bien variables intervenientes, inherentes al mismo proceso de la docencia y a su evaluación.

#### ***La investigación sobre los factores extraclase que afectan la efectividad docente***

La investigación producida durante los últimos veinte años sobre las variables ajenas al proceso de enseñanza-aprendizaje que afectan el desempeño del docente es muy extensa. Para llevar a cabo este trabajo se revisaron los índices ERIC y Psyclit. La búsqueda permitió agrupar los artículos en cuatro grandes categorías: 1) características del profesor, 2) características del alumno, 3) características del curso y la disciplina y 4) el uso de los resultados de los cuestionarios de evaluación con fines sumativos.

#### ***Características del profesor que afectan la evaluación***

Sin duda, las características del profesor tienen una marcada influencia en su desempeño docente. De hecho, la investigación (Marsh, 1980; 1984) indica que estos cuestionarios, más que evaluar el curso, califican al profesor. La investigación producida le ha dado marcada atención a los siguientes factores: a) personalidad, b) productividad investigativa, c) sexo o género, d) edad, e) experiencia y f) categoría académica.

#### ***La personalidad del profesor***

Aunque varios estudios indican que no existe relación entre el desempeño del profesor, medido a través de los cuestionarios de evaluación, y su personalidad (Abrami, Perry y Leventhal, 1982), la mayoría de las investigaciones concluyen que una parte sustancial de la efectividad docente depende de la personalidad del instructor. La investigación realizada durante las décadas de los cincuenta y sesenta (por ejemplo, Bending, 1955; Maslow y Zimmerman, 1956; Issacson, McKeachie y Milliholland, 1963) no produjo evidencias sustantivas sobre la relación entre la personalidad del profesor y su desempeño

docente (Murray, 1975). Como lo señalaron Costin y Grush (1973), la investigación de aquella época carecía de una clara definición de las variables investigadas. Fue hasta la década de los setenta y, especialmente la de los ochenta cuando la investigación produjo evidencias sustanciales de la relación existente entre personalidad y efectividad docente.

Uno de los primeros estudios exitosos fue el de Costin y Grush (1973). Estos investigadores encontraron que el comportamiento en el salón de clase tiene alta correlación con la descripción de los rasgos de personalidad del docente. Rasgos como: "juicioso, inspira confianza y comprensivo" fueron significativos en esas descripciones. En un estudio similar realizado en una universidad privada mexicana (García y Rugarcía, 1985) se le pidió a los alumnos de ingeniería de semestres avanzados que describieran, en términos de su personalidad y método de enseñanza, al mejor y al peor profesor que hubieran tenido durante sus estudios. La mayoría de las descripciones de los alumnos se centraron en los rasgos de personalidad de los profesores. El mejor profesor fue descrito como: "amable dentro y fuera del salón, de buen humor, alegre, sonriente, joven, adulto con espíritu juvenil, de mucho empuje". El peor profesor fue descrito como: "apático, pasivo, déspota, impuntual, poco interés por los alumnos, egoísta, se cree un sabelotodo".

Harry Murray, de la Universidad de Western Ontario, es uno de los investigadores más importantes de la relación entre personalidad y efectividad docente. Tal vez, junto con el mencionado estudio de Costin y Grush, una de las investigaciones más importantes de la década de los setenta fue el de Murray (1975). El autor empleó los cuestionarios de evaluación y un instrumento para que los pares calificaran la personalidad del profesor. El hallazgo del estudio indica que rasgos del profesor como liderazgo, extroversión, objetividad y bajo nivel de ansiedad predicen el 67% de la varianza en los cuestionarios de evaluación.

La investigación realizada durante los ochenta fue importante para ampliar y confirmar los hallazgos de la década anterior. Se confirmó que los rasgos de personalidad del profesor explican más del 50% de la varianza en la efectividad docente, medida a través de los cuestionarios de evaluación. Tomasco (1980) encontró que los profesores mejor evaluados son de dos tipos: uno, que es el más débil en términos estadísticos, es el que no autoritario y humilde; otro, el más fuerte y visible, es el que está orientado hacia los alumnos, es decir que les inspira confianza, respeta sus opiniones, es sensible a sus dificultades, es trabajador y amigable. Murray (1983) entrenó a observadores para que visitaran los salones de clase de profesores que habían obtenido altas, medias y bajas puntuaciones en los cuestionarios de evaluación. El autor concluyó que los mejores profesores son amigables, flexibles, pero al mismo tiempo tienen metas claras y toman la iniciativa en el salón de clase.

Erdle, Murray y Rushton (1985) concluyeron que el comportamiento en el aula está mediado por la relación entre personalidad y efectividad docente; el análisis de ruta (*path analysis*) confirmó que un 50%

de la varianza se explica por comportamientos del profesor que tienen que ver con su carisma y organización. Como en el estudio de Tomasco (1980), los maestros con altas puntuaciones en el cuestionario mostraron dos diferentes tipos de rasgos de personalidad; el primer tipo refleja orientación al logro (dominio, inteligencia y liderazgo) y el otro denota una orientación interpersonal (solidario, no autoritario y no defensivo).

En un extenso meta-análisis, Feldman (1986) encontró que los profesores y alumnos muestran acuerdo en que los maestros efectivos son: energéticos, empáticos y cálidos, ascendientes, muestran liderazgo, reflexivos, intelectual y culturalmente sensibles, alta autoestima, flexibles y adaptables, emocionalmente estables, sociables y gregarios, brillantes e inteligentes, responsables, conscientes, persistentes y ordenados.

Probablemente el trabajo más reciente e importante que se ha escrito, y que resume y amplía la investigación sobre personalidad y efectividad docente, es el de Murray, Rushton y Paunonen (1990). Los autores investigaron la relación entre la evaluación de la personalidad por pares y los resultados en los CEDA de la efectividad docente en seis tipos diferentes de cursos de psicología. Además de encontrar que la efectividad docente del profesor varía de acuerdo con el tipo de curso, los investigadores encontraron que la evaluación en los CEDA está fuertemente asociada con las de la personalidad que hicieron los pares o colegas. La regresión múltiple indicó que entre 40% y 70% de la varianza en la efectividad docente se explica con base en cinco rasgos de personalidad. Los que mostraron las correlaciones más altas fueron: liderazgo, extroversión, liberalidad, solidaridad, curiosidad intelectual y adaptabilidad. Otro hallazgo importante fue que los rasgos de personalidad que explican la efectividad docente son diferentes de acuerdo con el nivel del curso, mientras que en los de "pregrado" los rasgos de liberalidad y extroversión mostraron altas correlaciones con la efectividad; los de logro y resistencia fueron los que presentaron las más altas correlaciones en el nivel de posgrado.

La investigación ha confirmado que los rasgos de personalidad asociados con la efectividad docente son diferentes de los relacionados con la creatividad en la investigación. En dos estudios independientes Rushton, Murray y Paunonen (1987) llevados a cabo por medio de los cuestionarios de evaluación, evaluación de pares y autoevaluación de los participantes, encontraron que el profesor efectivo se describe como liberal, sociable, extrovertido, con liderazgo, no ansioso, inteligente y estéticamente sensible; en cambio el investigador creativo es ambicioso, resistente, dominante, con liderazgo, agresivo, independiente e insumiso. El único rasgo que comparten los investigadores y los docentes es el de liderazgo.

De los estudios revisados se puede inferir que los rasgos de personalidad más frecuentes en los docentes considerados efectivos son: liderazgo, extroversión, solidaridad, inteligencia y curiosidad intelectual.

### *La productividad en investigación*

Ha existido un largo debate en la educación superior sobre la relación entre investigación y docencia; aun cuando ambas son funciones sustantivas de las más prestigiadas instituciones, la investigación ha llegado a ocupar un lugar predominante, como medio para lograr y mantener el prestigio de muchas instituciones de ese nivel. El debate, tanto en el interior de las instituciones como en la investigación sobre efectividad docente, ha consistido en determinar si los investigadores más productivos son mejores profesores que aquellos que se dedican casi exclusivamente a las tareas docentes; al respecto ha habido argumentos en pro y en contra. Se dice que los investigadores pueden ser mejores profesores porque el dedicarse a la investigación los hace estar más actualizados en el campo del conocimiento. Asimismo, como se vio en la sección anterior, es probable que los rasgos de personalidad del investigador los haga imponer más carga de trabajo sobre sus alumnos, puesto que es más autoritario y demandante que el dedicado a la docencia.

La investigación realizada ha concluido, consistentemente, que sí existe una relación positiva, pero modesta, entre productividad en investigación y efectividad docente, y que la productividad no se opone a la efectividad docente. Centra (1983), en un estudio que incluyó una muestra de más de 4500 profesores de diversas instituciones estadounidenses, encontró que sólo en las ciencias sociales se presenta una relación positiva, pero discreta, entre el número de artículos publicados y la efectividad docente. Resultados similares se encontraron más recientemente (Noser, Manakyan y Tanner, 1996). De aquí se concluye que la productividad académica del profesor, en términos de investigación y publicaciones, no es necesariamente una condición para ser un buen docente.

Uno de los estudios más importantes es el de Feldman (1987); el autor realizó un meta-análisis de 43 investigaciones publicadas entre 1950 y 1983. Los resultados generales indican que existe, en promedio, un coeficiente de correlación significativo, pero pequeño, (.13) entre productividad académica y efectividad docente. Las dimensiones de la efectividad docente que correlacionaron significativamente, parecen las esperadas por los profesores dedicados a la investigación: conocimiento de la materia (.21), amplitud intelectual del profesor (.15) preparación y organización del curso (.19) y claridad de los objetivos y requerimientos del mismo (.18).

Feldman también analizó la relación entre productividad académica y otras variables inherentes al profesor; no se encontraron relaciones consistentes entre la productividad y la categoría académica y la edad del profesor; el autor observó relaciones débiles pero significativas con la categoría académica del profesor y su edad. En un estudio realizado en la Texas A & M University por Troy, Freidrich y Troy (1992) confirmó que la categoría académica no se relaciona con la productividad. No obstante otros estudios (Rushton, et al., 1987) indican que en la disciplina de la psicología, sí existe relación entre edad

y productividad; el profesor alcanza la cima de su productividad entre los 35-44 años y declina paulatinamente de 1.2 artículos por año en revistas arbitradas a .9.

El estudio citado de Feldman señala que existen diferencias entre productividad de acuerdo con la disciplina académica. Las disciplinas en las que se percibe un impacto positivo en la efectividad docente son: ciencias sociales y humanidades. En cambio, en las ciencias naturales tiene poco impacto; se presume que ello se deba a que el tipo de investigación en esta disciplina es más complejo y abstracto; el trabajo de Feldman no arrojó conclusiones definitivas. Sin embargo, de acuerdo con un estudio reciente, llevado a cabo en el Reino Unido y Estados Unidos, Patrick y Stanley (1998) encontraron que no siempre la relación entre la calidad en la docencia y la investigación es débil o negativa. Algunos departamentos académicos en ambos países, en especial de las universidades más antiguas, destacan en ambas funciones.

Diversas investigaciones concluyen que la productividad académica del profesor está relacionada con otras variables ajenas a la efectividad docente. Los estudios Clark y Centra (1985) y Krebs, Hurley y Smith (1991) sostienen que la productividad del profesor está relacionada con la calidad del programa doctoral. El estudio de Clark y Centra (1985) también halló que el número de años de haber obtenido el grado es un factor significativo en la productividad. Los profesores con cinco años de haberse graduado son más productivos que aquellos que obtuvieron el grado tres o 13 años después. Otro factor que afecta la productividad académica en las disciplinas económicas es el tamaño y las características de la institución universitaria. Los profesores que trabajan para instituciones que tienen más de 15 mil alumnos son más productivos que aquellos que lo hacen en las de menor tamaño (Noser, Manakyan y Tanner, 1996).

#### *Género del docente*

La investigación sobre efectividad docente y género ha cobrado mayor importancia en la medida en que se ha incrementado la proporción de docentes del sexo femenino que ingresan a las instituciones de educación superior. En general, el grueso de la investigación concluye que las profesoras y profesores son evaluados de manera diferente y que esta calificación también se ve afectada por el sexo de los estudiantes y la disciplina de que se trate. Aunque existe cierta inconsistencia, varios estudios han encontrado que las mujeres son igual o mejor evaluadas que los hombres.

Los trabajos más importantes, que sintetizan en gran parte el estado de conocimiento, fueron realizados por Feldman. En un primer meta-análisis de 17 estudios experimentales y de laboratorio (Feldman, 1992) concluyó que en las investigaciones analizadas no existen diferencias consistentes en el desempeño o efectividad docente entre profesores del sexo masculino y femenino. Asimismo, encontró que las disidencias entre las dimensiones de la docencia, de acuerdo con el género, no son consistentes.

Sin embargo, en una investigación posterior Feldman (1993) dedicada a analizar estudios realizados en el salón de clase, no confirmó los hallazgos del laboratorio. La evaluación global de la efectividad docente indicó que, aunque la correlación entre ésta y el género es muy pequeña como para ser tomada en consideración (.02), las profesoras son significativamente mejor evaluadas que los hombres en 28 de los 39 estudios analizados. En términos de las dimensiones de la efectividad docente las mujeres fueron significativamente mejor evaluadas en sensibilidad y preocupación con el nivel del grupo y su progreso, así como en su apertura a la opiniones de otros, disponibilidad y ayuda. Las dimensiones en que los hombres resultaron mejor valorados fueron la claridad y el conocimiento de la materia, pero su magnitud fue muy pequeña (.06 -.03). Goodwin y Stevens (1993) arribaron a resultados similares a los de Feldman. Asimismo señalan que los profesores le dan mayor importancia a los resultados de la evaluación de los alumnos que las profesoras, y que éstas son más propensas a buscar ayuda para mejorar su práctica docente que sus contrapartes del sexo masculino.

Los alumnos tienden a evaluar mejor a los profesores del mismo sexo que al sexo opuesto. De acuerdo con Kierstead, Agostino y Dill (1988), los docentes que encajan en los estereotipos sociales asignados reciben mejores evaluaciones; las mujeres son más negativamente evaluadas que los hombres si no tienen éxito en cumplir con las expectativas apropiadas correspondientes a su género.

En un estudio experimental más reciente Freeman (1994) aporta nuevos e interesantes hallazgos sobre la interacción entre género del profesor y del alumno. El autor encontró que la interacción entre género y evaluación se da por el rol que el profesor juega en el salón de clase. Los alumnos prefieren, y evalúan mejor a aquellos instructores que poseen tanto características masculinas como femeninas, independientemente de su género. Basow (1995) confirmó los hallazgos de Feldman (1993). Esta investigadora encontró que las evaluaciones de los profesores varones no se ven afectadas por el género de los alumnos, en cambio sí afectan a las profesoras; que tienden a recibir las más altas calificaciones de estudiantes del sexo femenino. Asimismo encontró que la interacción entre género y evaluación varía dependiendo de la disciplina. Los profesores de ciencias sociales son mejor valorados que las profesoras en entusiasmo, estimulación y retroalimentación. En humanidades las mujeres tienden a recibir mejores calificaciones en retroalimentación. En ciencias los profesores tienen evaluaciones globales más altas que las mujeres, pero no existe interacción entre el género del alumno y la evaluación del profesor.

#### *Categoría académica, edad y experiencia del profesor*

En general, la investigación reciente sobre efectividad docente ha dejado de prestar atención a las variables relacionadas con la categoría académica, edad y experiencia del profesor; a pesar de que se ha logrado identificar que no existen patrones consistentes de relación entre estos sesgos, los hallazgos de investigación no han logrado explicaciones suficientes sobre las causas de estos patrones y cómo influyen estas variables en la efectividad docente. Los estudios más comprensivos son el meta-análisis

de Feldman (1983) y la síntesis de Marsh (1987); estos trabajos, a menos que se indique otra cosa, han sido las fuentes de referencia para esta sección.

#### Categoría académica

Aunque se podría presumir que la categoría académica estaría asociada con la edad y experiencia del profesor, los hallazgos indican que esta variable es independiente de las otras mencionadas. El estudio de Feldman encontró que en dos terceras partes de las investigaciones revisadas no existe relación entre efectividad docente y categoría académica. Sin embargo, algunos estudios sí han notado una relación positiva, pero baja entre estas dos variables (entre .06 y .26).

Al analizar las dimensiones de la efectividad docente en los estudios que hallaron una relación significativa en la evaluación global, Feldman encontró que los profesores con alta categoría académica obtienen mejores evaluaciones en conocimiento de la materia, expansividad intelectual, así como en valor del curso que imparten. En cambio los docentes con baja categoría académica obtienen mejores evaluaciones en características relacionadas con su personalidad como: la promoción de la discusión y apertura hacia las opiniones de otros, justicia, preocupación por los alumnos, ayuda, disponibilidad, reto intelectual y promoción de ideas independientes en los alumnos.

#### Edad y experiencia docente

En general, no existen relaciones significativas entre edad y experiencia con la efectividad docente (al menos cuando se mide a través de los CEDA). Feldman indica que en los pocos estudios que han encontrado diferencias significativas, la relación es inversa y presenta la forma de una J invertida, es decir a menor experiencia y edad, mayor efectividad. Algunos de los trabajos revisados por este autor concluyen que la puntuación en los CEDA se incrementa por cada año de experiencia, y llega a su cima en el cuarto año y a partir de éste empieza a declinar.

Feldman analizó por separado las dimensiones de la efectividad docente de acuerdo con la edad y experiencia de los profesores. Los resultados indican que los jóvenes son mejor evaluados en aspectos tales como: entusiasmo por la materia, claridad, entendimiento y ayuda. En el caso de los profesores más experimentados, se da una relación inversa con las dimensiones de claridad, entusiasmo, apertura, promoción de preguntas y discusión y apertura hacia las opiniones de otros. Estos resultados son congruentes con lo que recientemente Renaud y Murray (1996) encontraron: los rasgos de personalidad varían de acuerdo con la edad del instructor y algunos de ellos declinan con la adultez; la efectividad docente está inversamente relacionada con la edad del instructor.

El meta-análisis de Feldman tampoco encontró, en la mayoría de los doce estudios revisados, relación entre el entrenamiento o capacitación docente que imparten algunas universidades a sus profesores y la efectividad docente. Sería interesante investigar esta relación en universidades latinoamericanas, ya que

se ha asumido que los cursos de didáctica o pedagogía influyen en la efectividad docente de los profesores. Feldman, con base en datos provenientes de su meta-análisis, cree que los cursos benefician a los profesores más preparados que aquellos con menos preparación. Dicho autor tampoco encontró relación significativa entre la experiencia en impartir un determinado curso y la efectividad docente.

#### Lenidad en la asignación de notas (calificaciones)

Entre los profesores universitarios no es extraño oír que un factor que influye en las evaluaciones de la instrucción por los alumnos es la lenidad o blandura del profesor al otorgar las notas parciales o finales del curso. Se dice que los instructores que son más blandos en la asignación de notas reciben mejores evaluaciones de los alumnos sobre su desempeño.

La conclusión general de las diversas investigaciones y del meta-análisis de Feldman (1976) es que la lenidad del maestro en la evaluación de sus alumnos o la nota que esperan obtener éstos sí afecta la evaluación del curso y del profesor. Sin embargo, los coeficientes de correlación entre estas dos variables son bajos (entre 0.10 y 0.30). Marsh (1980) encontró que, cuando se controla el interés previo por el curso, la nota esperada por el alumno sólo explica el 2.6% de la varianza en los CEDA; dimensiones como el valor en el aprendizaje, entusiasmo del instructor y evaluación global del curso explican cuatro veces más la varianza en las puntuaciones de efectividad del instructor. Por lo tanto, aunque la nota esperada por el alumno influye en la efectividad docente, su valor real no es tan alto y, por consiguiente, el sesgo es de menor cuantía.

En un estudio experimental (Abrami, Dickens, Perry y Leventhal, 1980) encontraron una relación significativa entre notas y puntajes en los CEDA, pero no para todos los tipos de cursos; algunos grupos en los que los profesores dieron notas finales bajas, obtuvieron evaluaciones más altas que en otros donde otorgaron notas mayores. Sin embargo, una investigación reciente en la Universidad de Washington (Greenwald y Gillmore, 1997) señala que la lenidad en las notas puede explicar hasta el 50% de la varianza en la efectividad docente, pero que el otorgar altas notas no garantiza necesariamente obtener altas evaluaciones aunque en el caso de que el mismo instructor imparta el mismo curso a dos grupos diferentes y sólo varíe en cada uno de ellos la manera de evaluar, en el curso que otorgue más altas notas el instructor obtendrá más altas evaluaciones por parte de los alumnos.

La alta correlación entre notas y evaluación de los cursos disminuye cuando se toma al grupo como unidad de análisis y no al alumno individualmente (Gigliotti y Buchtel, 1990). Otros hallazgos más específicos indican que, a pesar de que algunos instructores otorgan altas e inmerecidas notas a los alumnos, el nivel de satisfacción por el curso no se incrementa (Howard y Maxwell, 1982) ni la asistencia a clase por los estudiantes (Vasta y Sarmiento, 1979).

**Características del alumno***Interés previo por el curso y carga de trabajo*

La investigación (Marsh, 1980; 1983) ha encontrado que el interés previo por tomar el curso, sin importar si éste es obligatorio u optativo, por parte del alumno es uno de los factores extraclase que más impacta la efectividad docente. Como se señaló antes, esta variable explica un porcentaje mayor de la varianza (4%) que la nota o calificación esperada por el alumno (2.7%). El interés por el curso está estrechamente vinculado con la motivación del alumno. Los hallazgos de Marsh (1980; 1983) señalan que el interés por el curso tiene impacto en la evaluación global del mismo y en la dimensión de la efectividad relacionada con el valor de lo aprendido; contrariamente a lo que se esperaba, la carga de trabajo que implica el curso afecta positivamente la evaluación: a mayor carga de trabajo, mejor es la evaluación para el instructor.

**Factores asociados con las características de los cursos y las disciplinas***Nivel en que se imparte el curso*

Investigaciones realizadas en los setenta y mediados de los ochenta (Feldman, 1977; Erdle y Murray, 1986) no encontraron diferencias significativas entre la efectividad global del docente y el nivel del curso impartido, sea de los primeros semestres o de los últimos años de nivel de licenciatura (o grado equivalente) o de posgrado. Empero, hallazgos de investigaciones posteriores muestran que sí existen diferencias y éstas se refieren a dimensiones específicas de la efectividad docente. Smith y Cranton (1992) investigaron las diferencias en las percepciones de los estudiantes sobre las habilidades docentes de 2816 grupos de una facultad de Comercio canadiense, los autores encontraron que las dimensiones de la efectividad docente varían de acuerdo con el año en que se encuentran cursando los estudios universitarios y si los cursos son de posgrado. Estos autores indican que los mejores predictores de la efectividad docente global, independientemente del nivel o año de estudios, son el interés, la atmósfera y la evaluación. Esas dimensiones son predictoras de la efectividad en los cursos del tercer año del *bachelor* o licenciatura y en los cursos de posgrado; en los primeros años de universidad la organización y claridad son más significativas para explicar la efectividad que el resto de las dimensiones. Cranton y Smith señalan que la claridad y organización de la clase son más decisivas para alumnos más jóvenes y experimentados.

En relación con los cursos de posgrado, en una sección anterior (personalidad del profesor) se había señalado que los alumnos, dependiendo del nivel, *bachelor* o posgrado, valoran de diferente manera los rasgos de personalidad asociados con la efectividad docente Murray, *et al.*, (1990). Otro hallazgo interesante es que los alumnos de posgrado evalúan más alto a los profesores que los de pregrado o licenciatura (Marsh y Bailey, 1993); estos autores señalan que a pesar de estas diferencias en los niveles

educativos, el instructor juega un papel más importante en los resultados de la evaluación que el nivel del curso (licenciatura o posgrado).

#### *Tamaño del grupo de instrucción*

Aunque existen estudios que demuestran que los grupos pequeños otorgan evaluaciones de instrucción tan altas como los grupos grandes (Marsh, Overall y Kesler, 1979), el grueso de la investigación realizada indica que el tamaño del grupo influye moderadamente en la efectividad del instructor. En un meta-análisis de 52 investigaciones, Feldman (1984) encontró que la correlación promedio entre tamaño del grupo y efectividad docente, medida por los CEDA, es de -.09; la correlación es inversa, los grupos de menor tamaño de alumnos son mejor evaluados que los de mayor tamaño (Greenwald y Gillmore, 1997). En algunos de los estudios analizados Feldman (1984) halló que se da una relación curvilínea, en forma de *U*, entre esas dos variables; es decir, los grupos de mayor y menor tamaño reciben mejores evaluaciones que los de tamaño mediano. Algo semejante halló Marsh y colaboradores (Marsh, Overall y Kesler, 1979): los grupos grandes tienden a ser evaluados tan bien como los pequeños, pero en éstos califican mejor la dimensión relacionada con la calidad de interacción con el profesor que los grupos grandes. Asimismo, los resultados del estudio de Smith y Cranton (1992) con estudiantes canadienses señalan que las dimensiones de organización claridad y discusión varían significativamente de acuerdo con el tamaño del grupo.

#### *La naturaleza de la disciplina*

Muchas instituciones universitarias usan un solo cuestionario de evaluación y varias de ellas asumen, implícita o explícitamente, que la evaluación en un curso de ciencias es igual a la de otro en humanidades. Sin embargo, la investigación ha demostrado que la evaluación global y por dimensiones puede variar según la disciplina de que se trate. En general, y de manera consistente, los cursos y profesores de las disciplinas relacionadas con las ciencias y las ingenierías son evaluados por los alumnos con promedios más bajos que en las disciplinas relacionadas con el arte, humanidades y ciencias sociales. De ello podría desprenderse que el desempeño de los docentes no debe ser juzgado uniformemente sino tomando en consideración la naturaleza de la disciplina de donde se ubica el curso que imparte. Por ejemplo: un profesor del área de ingeniería, con un promedio de 3.3 en el cuestionario de evaluación de la docencia, no necesariamente es inferior a su contraparte de humanidades o lenguas extranjeras con un promedio de 4.3.

En otro de sus meta-análisis de 11 investigaciones Feldman (1978) concluyó que los cursos en inglés, humanidades, arte y en lenguas extranjeras se ubican en los rangos alto y medio-alto en la evaluación por parte de los estudiantes; ciencias sociales (ciencia política, sociología, psicología y economía) cae dentro del rango medio-bajo; las ciencias (con excepción de la biología) y las ingenierías caen dentro del rango bajo de evaluación. Cashin (1990) empleó los datos generados por la aplicación del instrumento diseñado por él y sus colaboradores, el Instructional Development and Efectiveness Assessment (IDEA) y

el Student Instructional Report (SIR) del Educational Testing Service; el autor arribó a conclusiones similares a las de Feldman (1978): las artes y las humanidades son las disciplinas (tanto los cursos como sus profesores) que los alumnos evalúan más alto; las disciplinas que son calificadas con promedios más bajos son negocios (cuantitativo), contaduría, ciencias e ingeniería, física y matemáticas. Las disciplinas que se ubican dentro del rango medio de evaluación son las ciencias sociales, de la salud y la biología. Cashin explica parte de estas diferencias por medio de la orientación cuantitativa *versus* cualitativa: entre más cuantitativo sea el curso o disciplina, más severa será la evaluación del profesor por los estudiantes, de ahí que los maestros de las ciencias e ingenierías sean evaluados con menores promedios que los de las otras disciplinas.

Otro grupo de investigaciones se ha dedicado a explicar las razones de las diferencias disciplinarias en la evaluación. Para ello han empleado el modelo de Biglan (1973a, 1973b) en el cual Kuhn clasifica las disciplinas bajo tres criterios: 1) con base en el desarrollo de sus paradigmas: paradigmáticas (duras)/preparadigmáticas (suaves); 2) con base en su aplicabilidad: puras/aplicadas; 3) y con base en su orientación con el sistema de vida: animadas/inanimadas (*life system nonlife system*). Por ejemplo, a las ingenierías civil, computacional y mecánica, Biglan las agrupa en duras (paradigmáticas), aplicadas e inanimadas; en cambio la antropología y la psicología son suaves (preparadigmáticas) puras, y animadas; la educación se ubica en aplicada, suave y animada.

Neumann y Neumann (1983), de la Universidad Ben Gurion de Israel, realizaron, probablemente, el primer estudio de diferencias disciplinarias en la evaluación docente basado en el modelo de Biglan. Los autores arribaron a las mismas conclusiones que las de estudios posteriores: existen diferencias en la evaluación de los alumnos entre las disciplinas blandas y duras. Barnes y Barnes (1993) condujeron un estudio para identificar las dimensiones de la efectividad docente asociadas con la clasificación de Biglan. Los autores encontraron que estas dimensiones varían de acuerdo con el tipo de disciplina. Una de las dimensiones que más discriminó a lo largo de los diferentes cursos fue la cobertura del conocimiento; las que menos se discriminaron fueron la asignación de notas (*grading*) e interacción con el grupo. Hallazgos más específicos indican que el entusiasmo del instructor es un constructo relacionado con las disciplinas duras/suaves, pero tiene mayor variabilidad con las puras/aplicadas. En este trabajo la cobertura del conocimiento no fue una dimensión útil para evaluar la instrucción en las disciplinas duras/suaves en relación con la evaluación del curso, pero sí en relación con la del instructor; la razón probable es que la cobertura del conocimiento en el curso está determinada por la naturaleza de la disciplina. La conclusión final del estudio es que un instrumento único de evaluación de la docencia no se puede aplicar uniformemente a todas las áreas. Los autores indican que en las disciplinas suaves o preparadigmáticas las habilidades docentes del profesor son más importantes que en las paradigmáticas debido a que las primeras tienen poco desarrollada la teoría o paradigmas.

En un estudio menos reciente Hearn (1985) llegó a conclusiones parecidas. El autor utilizó la tipología de Holland para clasificar las ocupaciones y encontró que la habilidad docente es más importante en las profesiones artísticas (artes y humanidades) que en las investigativas (biología o física) o realistas (ingenierías). Incluso, en disciplinas que pertenecen a la misma clasificación existen diferencias. Hativa (1996) condujo un estudio para determinar el perfil de los profesores de física y química que emerge de los cuestionarios de evaluación. El trabajo concluyó que, aunque los promedios de evaluación son semejantes, existen variaciones en algunas dimensiones como la presentación interesante y relación con los estudiantes.

Aunque parece una tarea difícil construir instrumentos de evaluación para cada grupo de disciplinas, las instituciones universitarias sí podrían tomar en consideración que no es adecuado evaluar uniformemente a los profesores y que deben tomarse en cuenta las diferencias entre las disciplinas e, incluso, programas académicos (Neumann y Neumann, 1983).

### **¿El uso de los resultados de los cuestionarios de evaluación con fines sumativos?**

Los resultados de los cuestionarios de evaluación pueden tener dos fines básicos: uno formativo, para proveer de retroalimentación al profesor sobre su desempeño y otro sumativo. Sobre el uso de los cuestionarios con propósitos formativos existe consenso. Se ha concluido que es un medio útil, pero modesto, para mejorar el desempeño en las tareas docente si se conjunta la retroalimentación con asesoría al docente (Centra, 1973; Marsh y Roche, 1993). Los aspectos más útiles para el instructor —en orden descendente— son información sobre la interacción con los estudiantes, la manera de evaluar a los alumnos, la evaluación global en el cuestionario y sobre los aspectos estructurales del curso (Schmelkin, Spencer y Gellman, 1997). Los hallazgos también indican que los profesores que voluntariamente evalúan sus cursos no obtienen promedios más altos que aquellos que no son evaluados voluntariamente (Cashin y Perrin, 1983).

Un tema controvertido es el uso de los resultados de los cuestionarios de evaluación con fines sumativos. En las universidades estadounidenses es frecuente que la valoración del desempeño docente se utilice para otorgar definitividad en el empleo, promoción o aumentos salariales y que sea realizada tomando en consideración los resultados de los cuestionarios de evaluación; alrededor de un 10% de las universidades y colegios de ese país hace público o permite el libre acceso de los estudiantes a los resultados de las evaluaciones de los cursos impartidos por los profesores (Haskell, 1997).

Si bien los estudios no cuestionan la validez de los cuestionarios de evaluación para medir la efectividad docente, sí interrogan o se muestran precavidos sobre su uso para tomar decisiones relacionadas con la definitividad, promoción o aumento salarial de los profesores. Los resultados de la evaluación no deben

ser la única fuente para decidir; sólo debe considerarse el promedio global y no el específico para cada dimensión del cuestionario; la decisión debe estar basada en el promedio general de los cursos que ha impartido el profesor y no en una minoría de ellos (McCallum, 1984; Abrami, 1989; McKeachie, 1997).

Se han realizado estudios que señalan que los profesores estadounidenses aceptan el uso de los cuestionarios de evaluación tanto para propósitos formativos como sumativos (por ejemplo, Schmelkin, Spencer y Gellman, 1997). Sin embargo, otros trabajos y ensayos advierten el peligro de que el uso de los cuestionarios contribuya al declive de la educación superior debido a la presión impuesta sobre los profesores, al ligar los resultados de la evaluación con su salario o permanencia en la institución; ello provoca baja moral y ablandamiento de las exigencias en relación con la carga de trabajo y asignación de notas a los estudiantes (Renner, 1981; Ryan, Anderson y Brichler, 1980). Probablemente el estudio más reciente, consultado y polémico sea el de Haskell (1997); este autor, basado en una extensa revisión bibliográfica, documenta los problemas señalados y advierte que el uso de los cuestionarios de evaluación con propósitos sumativos atenta contra la libertad académica y el derecho a la definitividad en el empleo de los profesores; por esas razones Haskell concluye que los cuestionarios de evaluación no son válidos. El artículo de Haskell ha tenido una ardiente respuesta de Theall (1997), el cual defiende su validez con propósitos tanto formativos como sumativos.

Si bien la polémica no está resuelta, puede concluirse que el uso de los cuestionarios con propósitos formativos es ampliamente aceptado, pero que tal uso con propósitos sumativos debe ser cauteloso. La experiencia de algunas instituciones mexicanas como la Universidad Iberoamericana y el Instituto Tecnológico de Estudios Superiores de Monterrey, indican que es más útil y prudente el uso de los cuestionarios como medio de retroalimentación que con fines de promoción o para asegurar la permanencia en el empleo de los profesores.

### **Resumen y discusión**

El propósito de este trabajo fue realizar una síntesis de la investigación sobre los cuestionarios de evaluación de la docencia que han estudiado los factores extraclasses o sesgos que afectan la efectividad docente. El estudio identificó que los factores o sesgos pueden agruparse en aquéllos concernientes al profesor y los que se refieren al curso y la disciplina; adicionalmente se analizaron los efectos que tiene el uso de los cuestionarios con fines sumativos (definitividad, promoción o aumento salarial).

En relación con las características del profesor, el efecto de su productividad en investigación, género, edad, experiencia y manera de asignar notas es mínimo. Sin embargo, las características de personalidad del profesor juegan un papel preponderante en la evaluación de su efectividad docente. En relación con las características de los cursos y la disciplina, el interés previo por el curso, nivel en que se imparte y tamaño del grupo también afectan la efectividad docente, pero su efecto es relativamente

menor. En este rubro el efecto mayor en la evaluación de los cursos tiene que ver con las características de la disciplina: las suaves son mejores evaluadas que las duras.

Si bien no se puede ignorar que los factores extraclase afectan la efectividad docente, diversas investigaciones han concluido que el rendimiento o aprendizaje que logra el alumno en el curso es la variable que más afecta la efectividad del maestro (Cohen, 1981; Broder y Dorfman, 1994), es decir, lo que más valoran los alumnos al evaluar el desempeño del profesor es lo aprendido en el curso. De ahí que se pueda concluir que los factores o las variables extraclase no deben ser ignorados como posibles sesgos en la efectividad docente, pero su influencia no es tan grande como se presume y, por lo tanto, los cuestionarios de evaluación se comportan como instrumentos válidos y confiables para valorar la docencia.

El fin principal de estos cuestionarios es más formativo que sumativo. No obstante, si no es posible evitar su uso sumativo, es recomendable que los resultados de estos instrumentos no tengan un peso determinante en la evaluación del docente ni sean la única fuente de calificación de su desempeño.

Sin embargo, su utilización con propósitos sumativos debe ser cautelosa e ir acompañada de otros indicadores de la efectividad docente. Tres recomendaciones prácticas para el diseño y empleo de los cuestionarios de evaluación podrían derivarse de este trabajo:

- 1) Los cuestionarios de evaluación deben incluir reactivos que califiquen las ganancias del aprendizaje y el valor del curso para los alumnos.
- 2) No debe uniformarse la valoración de las disciplinas; los promedios de evaluación pueden variar según se trate de disciplinas duras o suaves, de manera que los resultados de un curso deben analizarse en relación con otros de la misma disciplina y se debe evitar hacer generalizaciones entre cursos que correspondan a disciplinas diferentes.
- 3) El uso más recomendable de los cuestionarios de evaluación es el que tenga fines formativos; su utilización por las instituciones de educación superior con propósitos sumativos debe ser cauteloso e ir acompañado de otros indicadores de desempeño del profesor.

En México, es mucha la presión interna y externa por evaluar la docencia y poco el conocimiento y la investigación sobre el uso y el abuso de estos instrumentos. Uno de los abusos más notables puede darse en la relación entre resultados en los cuestionarios de evaluación y la promoción salarial a través de los programas de estímulos. Ello no sólo puede pervertir el proceso mismo de la evaluación docente y *ablandar* las exigencias del profesor con los alumnos sino también puede ser un factor que genere muchas injusticias e inconformidades. Los resultados arrojados por los cuestionario de evaluación deben

ser considerados en su conjunto, los de un curso o de un semestre no son suficientes para evaluar el desempeño docente.

Los responsables del diseño de la política de evaluación docente y de su operación deben estar conscientes de los factores extraclase o sesgos que influyen en los resultados de la evaluación y de que el uso sumativo de estos instrumentos en los programas de estímulos debe procurarse que sea el mínimo. La evaluación de la docencia tiene la finalidad intrínseca de retroalimentar al profesor sobre su desempeño más que premiarlo o castigarlo monetariamente por los resultados obtenidos. Las recompensas y castigos del desempeño docente son, por naturaleza, más intrínsecos que extrínsecos.

### **Referencias bibliográficas**

- Abrami, P. C. (1989). "How should use student ratings to evaluate teaching?", en *Research in Higher Education* (30), pp. 221-227.
- Abrami, P. C.; Dickens, W. J.; Perry, R. P. y Leventhal, L. (1980). "Do teachers standars fo assigning grades affect student evaluations of instruction?", en *Journal of Educational Psychology* (72), pp. 107-118.
- Abrami, P. C.; Leventhal, L. y Perry, R. P. (1982). "The relationship between student personality characteristics, teacher ratings, and student achievement", en *Journal of Educational Psychology* (74), pp. 111-125.
- Barnes, L. L. y Barnes, M. W. (1993). "Academic discipline and generalizability of student evaluations of instruction", en *Research in Higher Education* (34), pp. 135-149.
- Basow, S. A. (1995). "Student evaluations of college professors: When gender matters", en *Journal of Educational Psychology* ( 87), pp. 656-665.
- Bending, A. W. (1955). "Ability and personality characteristics of introductory psychology courses rated competent and empathetic by their students", en *Journal of Educational Research* (48), pp. 705-709.
- Biglan, A. (1973a). "The characteristics of subject matter in different academic areas", en *Journal of Applied Psychology* (57), pp. 195-203.
- Biglan, A. (1973b). "Relationships between subject matter characteristics and the structure and output of university departments", en *Journal of Applied Psychology* (57), pp. 204-213.
- Broder, J. M. y Dorfman, J. H. (1994). "Determinants of teaching quality: What's important to students?", en *Research in Higher Education* (35), pp. 235-249.
- Cashin, W. E. (1990). "Students do rate different academic fields differently", en M. Theall y J. Franklin (eds.), *Student ratings of instruction*, New Directions for Teaching and Learning (43), San Francisco: Jossey-Bass.
- Cashin, W. E. y Perrin, B. M. (1983). "Do college teachers who voluntarily have courses evaluated receive higher students ratings?", en *Journal of Educational Psychology* (75), pp. 595-602.
- Centra, J. A. (1973). "Effectiveness feedback in modifying college instruction", *Journal of Educational Psychology* (65), pp. 395-401.

- Centra, J. A. (1983). "Research productivity and teaching effectiveness", en *Research in Higher Education* (18), pp. 379-389.
- Centra, J. A. (1993). *Reflective faculty evaluation*, San Francisco: Jossey-Bass.
- Clark, M. J. y Centra, J. A. (1985). "Influences on the career accomplishments of Ph.D.'s", en *Research in Higher Education* (23), pp. 256-269.
- Cohen, P. A. (1981). "Student ratings of instruction and student achievement: A meta-analysis of multisector validity studies", en *Review of Educational Research* (51), pp. 281-309.
- Costin, F. y Grush, J. (1973). "Personality correlates of teacher-student behavior in the college classroom", en *Journal of Educational Psychology* (65), pp. 35-44.
- Erdle, S. y Murray, H. (1986). "Interfaculty differences in classroom teaching behaviors and their relationship to student instructional ratings", en *Research in Higher Education* (24), pp. 115-127.
- Erdle, S., Murray, H., & Rushton, J. P. (1985). "Personality, classroom behavior, and student ratings of college teaching effectiveness: a path analysis", en *Journal of Educational Psychology* (77), pp. 394-407.
- Feldman, K. A. (1976). "Grades and college students' evaluations of their courses and teachers", en *Research in Higher Education*, (4), pp. 69-111.
- Feldman, K. A. (1977). "Consistency and variability among college students in rating their teachers and courses: A review analysis", en *Research in Higher Education* (6), pp. 223-274.
- Feldman, K. A. (1978). "Course characteristics and college students' ratings of their teachers and courses: What we know and what we don't", en *Research in Higher Education* (9), pp. 199-242.
- Feldman, K. A. (1983). "Seniority and experience of college teachers as related to evaluations they receive from students", en *Research in Higher Education* (18), pp. 3-123.
- Feldman, K. A. (1984). "Class size and college students' evaluations of teachers and courses: A closer look", en *Research in Higher Education* (21), pp. 45-116.
- Feldman, K. A. (1986). "The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: a review and synthesis", en *Research in Higher Education* (24), pp. 139-213.
- Feldman, K. A. (1987). "Research productivity and scholarly accomplishments of college teachers as related to their instructional effectiveness", en *Research in Higher Education* (26), pp. 227-298.
- Feldman, K. A. (1990). "An afterword for "the association between student ratings of specific instructional dimensions and student achievement": Refining and extending the synthesis of data from multisector validity students", en *Research in Higher Education* (31), pp. 315-317.
- Feldman, K. A. (1992). "College students' views of male and female college teachers: Part I-evidence from the social laboratory experiments" en *Research in Higher Education* (33), pp. 317-375.
- Feldman, K. A. (1993). "College students' views of male and female teachers: Part II-evidence from students' evaluations of their classroom teachers", en *Research in Higher Education* (34), pp. 151-211.
- Freeman, H. R. (1994). "Student evaluations of college instructors: Effects of type of course taught, instructor gender role, and student gender", en *Journal of Educational Psychology* (86), pp. 627-630.

- García, J. M. & Rugarcía, A. (1985). "Perfil del maestro motivante y del desmotivante en las carreras de ingeniería" en *Boletín Didac* (13), México: Universidad Iberoamericana.
- Gigliotti, R. J. y Buchtel, F. S. (1990). "Attributional bias and course evaluations", en *Journal of Educational Psychology* (82), pp. 341-351.
- Goodwin, L. D. y Stevens, E. A. (1993). "The influence of gender on university faculty members' perceptions of "good" teaching", en *Journal of Higher Education* (64), pp. 166-185.
- Greenwald, A. G. y Gillmore, G. M. (1997). "Grading leniency is a removable contaminant of student ratings" en *American Psychologist* (52), pp. 1209-1217.
- Haskell, R. E. (1997). "Academic freedom, tenure, and student evaluation of faculty: Galloping polls in the 21st century" en *Educational Policy Analysis Archives*, núm. 5 (6) [revista electrónica]. Extraído de la Web, septiembre 30, 1997: <http://olam.ed.asu.edu/eppa/>.
- Hativa, N. (1996). "University instructors' ratings profiles: Stability over time, and disciplinary differences", en *Research in Higher Education* (37), pp. 341-365.
- Hearn, J. C. (1985). "Determinants of college students' overall evaluations of their academic programs", en *Research in Higher Education* (23), pp. 413-437.
- Howard, G. S. y Maxwell, S. E. (1982). "Do grades contaminate student evaluations of instruction?" en *Research in Higher Education* (16), pp. 175-188.
- Issacson, R. L., McKeachie, W. J. y Milholland, J. (1963). "Correlations of teacher personality variables and student ratings", en *Journal of Educational Psychology* (54), pp. 110-117.
- Kierstead, D. D'Agostino y Dill, H. (1988). "Sex role stereotyping of college professors: Bias in students' ratings of instructors", en *Journal of Educational Psychology* (80), pp. 342-344.
- Krebs, P. J., Hurley, R. B. y Smither, J. W. (1991). "Relationship of vocational personality and research training environment to the research productivity of counseling psychologists", en *Professional Psychology* (22), pp. 362-367.
- Marsh, H. W. (1980). "The influence of student, course, and instructor characteristics in evaluations of university teaching", en *American Educational Research Journal* (17), pp. 219-237.
- Marsh, H. W. (1983). "Multidimensional ratings of teaching effectiveness from different academic settings and their relation to student/course/instructor characteristics", en *Journal of Educational Psychology* (75), pp. 150-166.
- Marsh, H. W. (1984). "Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility", en *Journal of Educational Psychology* (76), pp. 707-754.
- Marsh, H. W. (1987). "Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research" [número completo], en *International Journal of Educational Research*, núm. 11 (3).
- Marsh, H. W. y Bailey, M. (1993). "Multidimensionality of students' evaluation of teaching effectiveness", en *Journal of Higher Education* (64), pp. 1-18.
- Marsh, H. W.; Overall, J. U. y Kesler, S. P. (1979). "Class size, students' evaluations, and instructional effectiveness", en *American Educational Research Journal* (16), pp. 57-69.

- Marsh, H. W. y Roche L. (1993). "The use of students' evaluations and individually structured intervention to enhance university teaching effectiveness", en *American Educational Research Journal* (30), pp. 217-251.
- Marsh, H. W. y Roche, L. A. (1997). "Making students' evaluation of teaching effectiveness effective", en *American Psychologist* (52), pp. 1187-1197.
- Maslow, H. A. y Zimmerman, W. (1956). "College teaching ability, scholarly activity, and personality", en *The Journal of Educational Psychology* (47), pp. 185-189.
- McCallum, L. W. (1984). "A meta-analysis of course evaluation data and its use in the tenure decision", en *Research in Higher Education* (21), pp. 150-158.
- McKeachie, W. J. (1997). "Student ratings. The validity issue", en *American Psychologist* (52), pp. 1218-1225.
- Murray, H. G. (1975). "Predicting student ratings of college teaching from peer ratings of personality types", en *Teaching of Psychology* (2), pp. 66-69.
- Murray, H.G. (1975). "Predicting student ratings of college teaching from peer ratings of personality types", en *Teaching of Psychology* (2), pp. 66-69.
- Murray, H. G. (1983). "Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness", en *Journal of Educational Psychology* (82), pp. 138-149.
- Murray, H. G.; Rushton, J. P. y Paunonen, S. V. (1990). "Teacher personality traits and students instructional ratings in six types of university courses", en *Journal of Educational Psychology* (82), pp. 250-261.
- Neumann, Y. y Neumann, L. (1983). "Characteristics of academic areas and students' evaluation of instruction", en *Research in Higher Education* (19), pp. 323-334.
- Noser, T. C.; Manakyan, H. y Tanner, J. R. (1996). "Research productivity and perceived teaching effectiveness: A survey of economics faculty", en *Research in Higher Education* (37), pp. 299-320.
- Patrick, W. J. y Stanley, E. (1998). "Teaching and research quality indicators and the shaping of higher education", en *Research in Higher Education* (39), pp. 19-41.
- Renner, R. (1981). "Comparing professors: How students ratings contribute to the decline in quality of higher education", en *Phi Delta Kappan*, núm. 63 (2), pp. 128-130.
- Rushton, J. P.; Murray, H. G. & Paunonen S. V. (1987). "Personality characteristics associated with high research productivity", en D. N. Jackson y J. P. Rushton (eds.), *Scientific excellence: Origins and assessment*. Newbury Park, California: Sage.
- Renaud, R. D. y Murray, H. A. (1996). "Aging, personality, and teaching effectiveness in academic psychologists", en *Research in Higher Education* (37), pp. 323-340.
- Ryan, J.; Anderson, J. A. y Birchler, A. B. (1980). "Student evaluation: The faculty responds", en *Research in Higher Education* (12), pp. 317-333.
- Schmelkin, L. P.; Spencer, K. J. y Gellman, E. S. (1997). "Faculty perspectives on course and teacher evaluations", en *Research in Higher Education* (38), pp. 575-592.
- Smith, R. A. y Cranton, P. A. (1992). "Students' perceptions of teaching skills and overall effectiveness across instructional settings", en *Research in Higher Education* (33), pp. 747-764.

- Theall, M. (1997). "On drawing reasonable conclusions about student ratings of instruction: a reply to Haskell and to Stake", en *Educational Policy Analysis Archives*, núm. 5, (8) [ revista electrónica ]. Extraído de la Web, septiembre 30, 1997: <http://olam.ed.asu.edu/eppa/>.
- Tomasco, A. T. (1980). "Student perceptions of instructional and personality characteristics of faculty: a canonical analysis", en *Teaching of Psychology* (7), pp. 79-82.
- Troy, M. E., Friedrich, K. y Troy, M. F. (1992). teaching or research? The relationship between scholarly productivity and students' judgments of teaching. *Trabajo presentado en la Reunión Anual de la American Evaluation Association*, Seattle.
- Vasta, R. y Sarmiento, R. F. (1979). Liberal grading improves evaluations but no performance. *Journal of Educational Psychology* (71), 207-211.

Recepción del artículo: 8 de diciembre de 1999.

Aceptado: 27 de junio de 2000.