

---

# ¿ES SESGADA LA EVALUACIÓN ESTUDIANTIL? EL CASO DE LA UNIVERSIDAD DE PUERTO RICO EN BAYAMÓN

---

*Horacio Matos-Díaz\**  
*Alfred J. Crouch Ruiz\*\**

La evaluación de la labor docente de los profesores es una práctica común en las universidades. En la Universidad de Puerto Rico en Bayamón (UPR-B), y en el resto del sistema de la UPR, son evaluados por los pares que integran el comité de personal departamental, por el director del departamento y por los estudiantes<sup>1</sup>. La evaluación de los estudiantes se realiza a finales de cada semestre académico mediante un cuestionario anónimo que contestan voluntariamente. Los resultados de los tres tipos de evaluación sirven como criterios para tomar decisiones importantes, como la permanencia y los ascensos.

\* Doctor en Economía, profesor del Departamento de Administración de Empresas de la Universidad de Puerto Rico, Bayamón, Puerto Rico, [hmatos@uprb.edu].

\*\* Magíster en Física, profesor jubilado del Departamento de Física de la Universidad de Puerto Rico, Bayamón, Puerto Rico, [ajcrouch@onelinkpr.net]. Este estudio se completó en el año académico 2006-2007 cuando el primero de los autores disfrutaba de una licencia sabática para investigación. Estamos en deuda de gratitud con nuestros colegas y amigos Gilberto Calderón y Dwight García. El primero acopió y documentó la información utilizada en el estudio; el segundo revisó y editó el manuscrito original. También agradecemos al doctor Andrés Rodríguez Rubio, anterior Rector de la UPR-B, sin cuyo apoyo no habríamos completado este trabajo. Las recomendaciones y sugerencias de dos árbitros anónimos de esta Revista ayudaron a mejorar el contenido: para ellos nuestro agradecimiento. Es de rigor señalar que los errores y omisiones que persistan son de nuestra única y total responsabilidad. Fecha de recepción: 27 de mayo de 2007, fecha de modificación: 3 de septiembre de 2007, fecha de aceptación: 3 de marzo de 2008.

<sup>1</sup> Este no parece ser el caso en los departamentos de economía de las universidades de Estados Unidos. De acuerdo con Becker y Watts (1999) sólo un 37% de los departamentos de economía de las universidades dedicadas a la investigación utiliza el método de evaluación por pares; mientras que un 52% de los departamentos de las universidades dedicadas a la enseñanza utiliza este método.

Se puede conjeturar que la evaluación estudiantil depende de la utilidad que obtienen los estudiantes (Anderson y Siegfried, 1997). Si esta utilidad dependiera exclusivamente de cuánto han aprendido, la evaluación cumpliría su cometido. Pero si en la percepción de la utilidad intervienen otros factores, esta evaluación se puede convertir en un concurso de popularidad. Entre esos factores se podrían incluir los siguientes: género, edad, capacidad intelectual, madurez emocional, grado de responsabilidad, año que cursan, estado civil, nivel socioeconómico, etc., de los estudiantes; rigor en la evaluación, experiencia, edad, género, grado y rango académico, motivación, entusiasmo, claridad de expresión, dotes histriónicas, características afectivas, etc., de los profesores; la naturaleza y el nivel de dificultad del curso, si es requisito o electivo, el tamaño de la clase, hora y días en que se dicta, si requiere laboratorios, número de créditos, etc.

El grado en que estas variables (algunas fuera del control del profesor y otras sujetas a su manipulación) inciden en la utilidad que perciben los estudiantes, y en su evaluación, puede viciar y desvirtuar el proceso<sup>2</sup>. La evidencia estadística tiende a confirmar que todas estas variables explican, en mayor o menor grado, parte de la variación observada en las evaluaciones estudiantiles<sup>3</sup>.

Cabe señalar que al institucionalizar la evaluación estudiantil se legitima el rol de supervisión de los estudiantes. De entes pasivos pasan a ser entes activos con poder para negociar las reglas con que sus ejecutorias serán medidas, es decir, los supervisados se transforman en supervisores. Sería ingenuo ignorar que los estudiantes y los profesores son conscientes de su mutua interdependencia. Los primeros saben que la permanencia y los ascensos de los segundos dependen de sus evaluaciones; mientras que los profesores son conscientes de

<sup>2</sup> En principio, una correlación positiva y significativa entre la “claridad de expresión” del profesor y su evaluación no constituye sesgo alguno pues este atributo contribuye al aprendizaje de los estudiantes. Pero un discurso claro, lúcido y elegante puede ser totalmente irrelevante y disparatado desde el punto de vista del objetivo académico. El ejemplo clásico es el “efecto del Dr. Fox”. Un actor profesional dicta una conferencia a un grupo de estudiantes universitarios. Una vez finalizada, los estudiantes hacen una evaluación muy favorable, basados en su claridad, elocuencia y dotes histriónicas. Pero la conferencia estuvo plagada de neologismos, términos sin sentido, contradicciones y conclusiones ilógicas (ver Nerger, Viney y Riedel, 1997, y su fuente original). De ahí la insistencia en atar la evaluación estudiantil al logro de las metas académicas.

<sup>3</sup> Para una muestra de las investigaciones sobre el tema publicadas fuera de la literatura económica, ver Centra y Gaubatz (2000), Nerger, Viney y Riedel (1997) y Stringer e Irwing (1998), así como las fuentes que citan.

que los estudiantes necesitan buenas calificaciones para progresar en sus estudios.

Amparados en la libertad de cátedra, los profesores definen unilateralmente las normas de evaluación y las “reglas del juego” que imperan en el salón de clases. Además, pueden aprender sobre la marcha cuáles son las pautas de comportamiento y las consideraciones afectivas que maximizan la utilidad percibida por sus estudiantes. Estos, a su vez, protegidos por el anonimato de la evaluación, pueden “castigar” a los profesores que les dan una baja calificación (a la que perciben como una baja “utilidad”) independientemente de la eficacia del profesor (“mayor utilidad” no implica necesariamente más conocimientos).

¿Por qué ambos bandos descartarían la posibilidad de una negociación que defina las condiciones del entorno? En la medida en que los profesores y los estudiantes se involucren solapada o expresamente en esa negociación se da al traste con la excelencia académica que busca (o debe buscar) la universidad. Es decir, si los estudiantes pueden “comprar” buenas calificaciones y los profesores evaluaciones, harán el menor esfuerzo académico posible. En ese caso no se lograría el aprendizaje esperado.

Infortunadamente, nadie ha definido qué se pretende medir con la evaluación estudiantil, ni qué criterios se deben satisfacer para obtener una “buena evaluación”. Es preocupante que aun en esas condiciones se haya institucionalizado el proceso, y todo indica que llegó para quedarse. Este estudio intenta arrojar luz sobre los posibles sesgos que pueden invalidar la evaluación estudiantil. Además, pretende sentar unos cimientos analíticos sobre los cuales se pueda comenzar una discusión rigurosa sobre el tema.

En la siguiente sección se discute la literatura relevante. Luego se define el objetivo del estudio, se presenta la información, y se discute y justifica el modelo de estimación. Después se presentan, interpretan y resumen los resultados.

## LITERATURA RELEVANTE

De acuerdo con Algozzine et al. (2004), el precursor del proceso de evaluación estudiantil fue Herman Remmers, quien elaboró la *Purdue Rating Scale for Instructors* en 1927, cuya validez y adecuación se empezaron a cuestionar prácticamente desde su fase experimental, pues, como señalan Sailor, Worthen y Shin (1997), el origen de la discusión se remonta a 1928. Desde entonces se ha documentado la correlación positiva y estadísticamente significativa entre la evalua-

ción estudiantil y las calificaciones que los estudiantes obtienen (o esperan) en los cursos respectivos, un resultado estadístico sobre el que hay consenso general, aunque continúa la controversia acerca de su interpretación (Simpson y Siguaw, 2000).

Algunos investigadores argumentan que esa correlación es una evidencia a favor del proceso de evaluación estudiantil puesto que cuando los estudiantes aprenden más y obtienen buenas calificaciones evalúan bien a sus profesores (Cohen, 1986; Marsh, 1987, y Moore, 2006). Otros investigadores la interpretan como evidencia de que la evaluación está viciada en alguna medida como resultado del proceso de negociación antes descrito (Hamilton, 1980, y Wachtel, 1998). Germain y Scandura (2005) resumen la controversia señalando que las conclusiones sobre las evaluaciones estudiantiles varían desde “válidas, confiables y útiles” hasta “inválidas, no confiables y carentes de utilidad”. Las críticas de Emery, Kramer y Tian (2003) van aún más lejos al señalar que el anonimato y la incapacidad de los estudiantes para evaluar objetivamente a sus profesores son elementos que servirían de base para procesos legales sobre difamación.

En la literatura económica, la discusión sobre las evaluaciones estudiantiles toma prominencia en el contexto del debate sobre la inflación de las calificaciones (Johnson, 2003, y Matos-Díaz, 2006), aunque otros autores tratan el tema motivados por consideraciones que no tienen relación alguna con ese debate. Por ejemplo, Anderson y Siegfried (1997) exploran la relación entre las evaluaciones y el género con datos de las evaluaciones de 189 cursos de introducción a la economía en 53 universidades de Estados Unidos y las calificaciones obtenidas en el “Test of Understanding College Economics” (TUCE II) de 1990. Una vez controladas ciertas características de los profesores y de los estudiantes, no detectan diferencias significativas por género en los cursos de macroeconomía. Pero encuentran que las profesoras tienden a recibir evaluaciones significativamente mejores que los profesores en los cursos de microeconomía.

Dos estudios adicionales utilizan la tercera edición de esa prueba (TUCE III) para abordar dos problemas diferentes. Finegan y Siegfried (2000) analizan el efecto de que el inglés no sea la lengua vernácula del profesor que dicta el curso. Mediante una escala de 1 a 5 que mide la evaluación estudiantil documentan la existencia de una prima de 0,4 puntos a favor de los profesores cuya lengua vernácula es el inglés. Bosshardt y Watts (2001) comparan las evaluaciones que los profesores hacen de sí mismos con las de sus estudiantes en cuatro aspectos: entusiasmo, preparación, habilidad para hablar inglés y

rigor en las calificaciones; y descubren que la autoevaluación de los profesores cuya lengua materna es el inglés y la de los demás sólo difieren en la “habilidad para hablar inglés”. Los estudiantes, por su parte, evalúan mejor en todos los aspectos a los profesores cuya lengua vernácula es el inglés.

McPherson (2006) utiliza datos de las evaluaciones de los estudiantes de economía realizadas en la Universidad del Norte de Texas durante 17 semestres (de 1994 a 2002) para analizar su relación con las calificaciones que esperaban en los cursos respectivos, y encuentra que las evaluaciones están significativa y directamente relacionadas con las calificaciones esperadas. De acuerdo con el autor, este resultado indica que los profesores pueden “comprar” buenas evaluaciones si les dan altas calificaciones a sus estudiantes.

Los tres primeros estudios mencionados se limitan a los cursos de introducción a la microeconomía y a la macroeconomía; sólo el cuarto considera otros cursos de pregrado en economía. Nuestra investigación amplía el ámbito de la discusión pues, además de los cursos de economía y de estadística, cubre otras 16 disciplinas que no han sido analizadas en la literatura. Además, al clasificar a los profesores por disciplinas, rangos y grados académicos; considerar la hora, el semestre y el número de estudiantes presentes en la evaluación; el género y las interacciones entre las variables, se puede estimar con mayor precisión el impacto de los efectos de cada variable sobre la evaluación. Los estudios anteriores no tienen semejante nivel de especificidad. Y esa es la contribución principal de este artículo.

Por otra parte, en este estudio se reconoce explícitamente la posibilidad de que la calificación que esperan los estudiantes sea una variable endógena y, por tanto, que esté correlacionada con los residuos de la ecuación que se utiliza para estimar la evaluación que recibirán sus profesores. La prueba de Hausman (1978) corrobora esta conjetura y se emplea el método de mínimos cuadrados en dos etapas (MC2E) en vez del método de mínimos cuadrados ordinarios (MCO).

## **OBJETIVO DEL ESTUDIO**

A diferencia de las universidades de Estados Unidos, donde se han discutido ampliamente los problemas y las limitaciones de las evaluaciones estudiantiles, el tema nunca se ha estudiado en la UPR-B (y probablemente en el resto del sistema de la UPR). Peor aún, no se han definido conceptual ni operativamente los criterios que deben regir el proceso ni, por tanto, el contenido y el alcance de las preguntas in-

cluidas en el cuestionario de evaluación. Tampoco se ha determinado qué aspectos de la información sobre los estudiantes son relevantes ni cómo obtenerlos manteniendo el anonimato. Además, no se ha justificado el carácter anónimo del proceso.

Este estudio pretende llenar el vacío conceptual y analítico que existe en la discusión local sobre el tema. Su propósito fundamental es determinar hasta qué punto y de qué manera las características de los profesores, de los cursos y de los estudiantes inciden en los resultados de las evaluaciones estudiantiles. Presta particular atención al efecto del género de los profesores y a su relación con los cursos que dictan, su rango y su grado académico. Para ello, se estudian las evaluaciones de 187 profesores permanentes o con contratos de prueba realizadas en 8 semestres de los períodos 1998-1999 y 2003-2004.

#### DESCRIPCIÓN DE LOS DATOS

Para la evaluación estudiantil la UPR-B utiliza un cuestionario anónimo, de marca sensible, que responden los estudiantes presentes en el salón de clases el día señalado. El cuestionario incluye preguntas sobre los siguientes aspectos: nombre del profesor, departamento, curso y hora de la evaluación; semestre y año académico; calificación que espera el estudiante en el curso; índice académico general del estudiante (IAG); disponibilidad y entusiasmo con que tomaría otro curso con el profesor que está evaluando. La información sobre los grados y rangos académicos de los profesores se actualiza cada semestre.

El cuestionario contiene 25 preguntas cuyas respuestas siguen la siguiente escala: totalmente de acuerdo = 4; moderadamente de acuerdo = 3; moderadamente en desacuerdo = 2; totalmente en desacuerdo = 1; no aplica = 0. La variable dependiente se define con base en las respuestas de los estudiantes.

El primer grupo de evaluaciones se hizo el segundo semestre del año académico 1998-1999. No están disponibles las del año académico 1999-2000 ni las del primer semestre de 2000-2001. Las siguientes se realizaron entre el segundo semestre del año académico 2000-2001 y el segundo semestre de 2003-2004. Se incluyen todas las evaluaciones de cada profesor realizadas en esos 8 semestres.

Las variables que se utilizan en el estudio se clasifican en tres grupos. El primero incluye las variables que describen las características de los estudiantes: índice académico acumulado (IAA) y calificación esperada en el curso. El segundo incluye las variables que definen las características de los profesores: género, edad, grado y rango académ-

mico; departamento al que está adscrito y curso en el cual se evalúa. El último describe algunas características del curso: naturaleza; hora, semestre y año en que se dictó; así como el número de estudiantes que hicieron la evaluación.

Cuadro 1  
Descripción de las variables

Variable	Descripción	Media	DE	Máx.	Mín.
Instructor	1, si es Instructor	0,2136	0,4099	1	0
Catedrático auxiliar	1, si es Catedrático Auxiliar	0,3564	0,4790	1	0
Catedrático asociado	1, si es Catedrático Asociado	0,3521	0,4776	1	0
Catedrático	1, si es Catedrático	0,0778	0,2679	1	0
Bachillerato o maestría	1, si no es doctorado	0,6805	0,4663	1	0
Doctorado	1, si es doctorado	0,3195	0,4663	1	0
Femenino	1, si es mujer	0,5539	0,4971	1	0
Edad	Años	45,21	8,3810	73	26
Edad <sup>2</sup>	Años al cuadrado	2.113,76	777,91	5329	676
Matemáticas	1, si es matemáticas	0,0690	0,2535	1	0
Física	1, si es de física	0,0424	0,2016	1	0
Química	1, si es química	0,0470	0,2116	1	0
Biología	1, si es biología	0,0702	0,2555	1	0
Ciencias de la Computación	1, si es computación	0,0673	0,2505	1	0
Electrónica	1, si es electrónica	0,0506	0,2193	1	0
Tecnologías de Ingeniería	1, si es ingeniería	0,0303	0,1714	1	0
Economía y Estadística	1, si es economía o estadística	0,0145	0,1196	1	0
Administración de Empresas	1, si el curso es de ADEM	0,1076	0,3098	1	0
Inglés	1, si es inglés	0,0957	0,2941	1	0
Español	1, si es español	0,0772	0,2669	1	0
Ciencias Sociales	1, si es ciencias sociales	0,0916	0,2884	1	0
Humanidades	1, si es humanidades	0,0484	0,2146	1	0
Educación	1, si es educación	0,1037	0,3048	1	0
Educación Física	1, si es educación física	0,0359	0,1860	1	0
Sistemas de Oficina	1, si es sistemas de oficina	0,0488	0,2154	1	0
Evaluación por la mañana	1, si se hizo en la mañana	0,6346	0,4816	1	0
Estudiantes presentes	N. <sup>o</sup> de estudiantes que evaluó al profesor	26,27	21,7457	175	1
ÍAA 1	1, si el ÍAG < 2	0,0865	0,2810	1	0
ÍAA 2	1, si el ÍAG = 2	0,0690	0,2535	1	0
ÍAA 3	1, si (2 < ÍAG < 3)	0,2566	0,4368	1	0
ÍAA 4	1, si el ÍAG = 3	0,1882	0,3909	1	0
ÍAA 5	1, si el ÍAG > 3	0,3997	0,4898	1	0
Calificación esperada	A = 4, ..., F = 0	2,7856	1,2807	4	0
(Calificación esperada) <sup>2</sup>	Variable al cuadrado	9,3997	5,8634	16	0
Semestre 2 (1998-1999)	1, si corresponde	0,1954	0,3965	1	0
Semestre 2 (2000-2001)	1, si corresponde	0,0323	0,1768	1	0
Semestre 1 (2001-2002)	1, si corresponde	0,1126	0,3161	1	0
Semestre 2 (2001-2002)	1, si corresponde	0,0941	0,2919	1	0
Semestre 1 (2002-2003)	1, si corresponde	0,1116	0,3149	1	0
Semestre 2 (2002-2003)	1, si corresponde	0,1068	0,3088	1	0
Semestre 1 (2003-2004)	1, si corresponde	0,1596	0,3662	1	0
Semestre 2 (2003-2004)	1, si corresponde	0,1877	0,3905	1	0
Variable dependiente:					
SET <sub>ij</sub>	Puntuación promedio	363,74	45,5737	400	100

El efecto de variables cualitativas como el género, el curso, el grado y el rango académico se modela utilizando variables binarias (0, 1) que indican la presencia o ausencia de la característica. Para las variables cardinales cuyo recorrido es muy limitado, como el índice

académico del estudiante, la hora en que se dictó el curso, y el semestre en que se hizo la evaluación, también se utiliza este artificio. Las variables continuas, como la edad del profesor y la calificación que espera el estudiante, se incluyen en forma normalizada<sup>4</sup>. En el cuadro 1 se describen estas variables y las que se mencionaron anteriormente.

#### BREVE DIGRESIÓN SOBRE LA ENDOGENEIDAD Y EL MÉTODO DE MC2E

En la literatura se discute el problema de la simultaneidad entre la evaluación que hace el estudiante y la calificación que espera en el curso (McPherson, 2006). Es decir, la calificación que espera el estudiante  $i$  en el curso que dicta el profesor  $j$  ( $C_{ij}$ ), lejos de ser una variable exógena y predeterminada, es endógena (aleatoria) y está correlacionada con los residuos de la ecuación de la puntuación que le atribuirá al profesor  $j$  ( $SET_{ij}$ ). Para ilustrar el problema, consideremos la ecuación  $SET_{ij} = f(Z_1, Z_2, \dots, Z_n, C_{ij} = g(\cdot))$ . Las  $Z$  son variables exógenas, predeterminadas e invariantes entre una muestra y otra, pero  $C_{ij}$  no lo es. Si hay variables endógenas en el lado derecho de la ecuación, el método de MCO arroja estimadores sesgados e inconsistentes. En esas circunstancias se debe utilizar el método de MC2E.

Cuando se adopta este método hay que encontrar una o más variables denominadas “instrumentos” que estén significativamente correlacionadas con  $C_{ij}$ , pero no con los residuos de la ecuación  $SET_{ij}$ . Si existen esos “instrumentos” se puede utilizar una prueba sugerida por Hausman (1978) para constatar si  $C_{ij}$  es endógena, que consiste en estimar dos conjuntos de ecuaciones de regresión con el método de MCO. En el primero se estiman  $C_{ij}$  y  $C_{ij}^2$  como función de las variables independientes que se utilizarán en  $SET_{ij}$ , más los “instrumentos”. En el segundo se estima  $SET_{ij}$  en función de las variables independientes (excluyendo a los “instrumentos”) más los residuos de las dos primeras ecuaciones. Los coeficientes de regresión de ambos residuos  $R_1$  y  $R_2$  se someten a una prueba de Wald para constatar la validez de la hipótesis nula:  $H_0 = \beta_{R_1} = \beta_{R_2} = 0$ .

<sup>4</sup> La transformación  $Z = (X - \mu)/\sigma$  permite expresar todas las variables continuas en la misma métrica: desviaciones estándar con respecto al promedio. Por tanto, la importancia relativa de las variables depende del valor del coeficiente de la regresión. Puesto que es una transformación lineal no altera el signo ni la significación estadística de las variables independientes.



El rechazo de tal hipótesis lleva a concluir que  $C_{ij}$  es endógena y que se debe adoptar el método de MC2E.

En la prueba de Hausman se utilizó como “instrumento” la calidad del estudiante. La mejor aproximación disponible de la “calidad de los estudiantes” son los índices académicos acumulados (ÍAA 1, ..., ÍAA 5) suministrados por los estudiantes cuando completaron la evaluación del profesor. A grandes rasgos, el método de MC2E se puede resumir como sigue. La relación que se quiere estimar es  $SET_{ij} = f(Z_1, Z_2, \dots, Z_n, \hat{C}_{ij}, \hat{C}_{ij}^2)$ . El conjunto de instrumentos es  $Z_1, Z_2, \dots, Z_n, \hat{ÍAA} 2, \dots, \hat{ÍAA} 5$ . En la primera etapa se estima la ecuación de regresión con el método de MCO para cada una de las variables independientes incluidas en  $SET_{ij}$  en función de los instrumentos, y se calcula su predicción ( $\wedge$ ). En la segunda etapa se sustituye cada una de ellas por su predicción y se estima el siguiente modelo utilizando MCO:  $SET_{ij} = f(\hat{Z}_1, \hat{Z}_2, \dots, \hat{Z}_n, \hat{C}_{ij}^E, \hat{C}_{ij}^2)$ . El vector de los coeficientes  $\hat{\beta}_{MC2E}$  se presenta en el cuadro 2.

Cuadro 2

Estimación de MC2E para explicar la evaluación del estudiante i al profesor j

Variables	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Constante	357,4853* (0,8824)	357,8742* (1,1734)	375,5782* (1,7445)	340,7773* (2,7123)	340,5208* (3,7959)
Instructor	4,6520* (1,0150)	3,7715* (1,0114)	2,8966* (1,0184)	3,7194* (1,4417)	4,8841** (2,0467)
Catedrático asociado	3,9101* (0,8894)	3,6149* (0,8847)	3,7745* (0,8935)	6,5570* (1,2712)	8,6834* (2,1496)
Catedrático	-0,0555 (1,4485)	-0,1747 (1,4481)	-1,1074 (1,4725)	-3,6698† (2,0847)	-1,4313 (4,0630)
Doctorado	4,1059* (0,7837)	4,5698* (0,7813)	-0,0275 (0,8868)	-5,8599* (1,2599)	-4,8120** (2,1077)
Edad	9,6215* (2,7951)	10,0103* (2,7793)	0,4881 (2,8553)	9,8037** (4,0260)	4,2642 (4,4280)
(Edad) <sup>2</sup>	-11,1045* (2,7490)	-11,4897* (2,7336)	-4,3527 (2,7964)	-13,0563* (3,9414)	-7,6664† (4,3186)
Femenino (F)	5,2894* (0,7270)	5,4496* (0,7294)	-3,0074* (0,8662)	-6,8760* (1,2310)	-12,6306* (4,2792)
Estudiantes presentes	2,4720* (0,3459)	2,4091* (0,3858)	1,0879* (0,4089)	2,1340* (0,5782)	1,9261* (0,6047)
Hora de la evaluación	0,0205 (0,7264)	-0,2394 (0,7268)	1,4544** (0,7405)	-1,1998 (1,0478)	-0,1154 (1,1067)
Biología			-20,2433* (1,7066)	3,4650 (2,5871)	6,7978† (4,0721)
Ciencias de la computación			-25,0776* (1,8249)	-28,3565* (2,5990)	-34,8962* (3,7211)
Electrónica			-38,8849* (1,9754)	-17,7245* (3,0068)	-23,9944* (3,9734)
Tecnologías e ingenierías			-41,3294* (2,3193)	-18,9682* (3,3828)	-19,6387* (4,5273)
Economía y estadística			-5,4016† (3,0115)	15,9392* (4,4295)	20,8584* (6,0735)
Administración de empresas			-19,3457* (1,5934)	-15,0309* (2,2565)	-26,1784* (3,8729)
Inglés			-2,5684 (1,6640)	0,4229 (2,3472)	-6,6943† (4,0399)
Matemática			-22,9787* (1,8025)	15,4222* (2,9110)	18,0613* (4,1890)

Física	-19,1489*	-3,5665	-8,5392**
	(2,0598)	(2,9541)	(3,5784)
Química	-15,5289*	4,2215	0,9208
	(1,9534)	(2,9764)	(8,8748)
Español	-6,3686*	-8,8176*	-30,3298*
	(1,6951)	(2,3874)	(5,1915)
Humanidades	-6,0125*	1,3983	-5,1228
	(1,9452)	(2,7483)	(4,0819)
Educación	-4,6641*	-8,0281*	6,3342
	(1,6183)	(2,2830)	(6,8119)
Educación física	-2,8334	-17,2401*	-21,4502*
	(2,1782)	(3,1374)	(4,2232)
Sistemas de oficina	5,6661*	16,7823*	20,8088*
	(1,9826)	(2,8343)	(3,1518)
Calificación esperada (CE)		25,8682*	34,8921*
		(8,8888)	(9,4099)
(CE) <sup>2</sup>		53,6156*	49,0345*
		(7,8027)	(8,2717)
F*Instructor			-3,2818
			(2,8918)
F*Catedrático asociado			-3,2328
			(2,6040)
F*Catedrático			-3,4190
			(4,6607)
F*Doctorado			0,0227
			(2,8116)
F*Matemática			-12,5890**
			(5,6439)
F*Física			-7,3637
			(17,4252)
F*Química			5,7732
			(9,3187)
F*Biología			-7,2003
			(5,3393)
F*Ciencias de la computación			21,2167*
			(6,4203)
F*Electrónica			15,8834**
			(6,5880)
F*Tecnologías e ingenierías			-4,1620
			(7,7883)
F*Economía y estadística			-13,2829
			(9,1017)
F*Administración de empresas			21,2007*
			(4,8791)
F*Inglés			12,3531**
			(5,0965)
F*Español			28,2434*
			(6,0262)
F*Humanidades			13,1932**
			(5,7674)
F*Educación			-11,6273
			(7,2713)
F*Educación física			12,4696†
			(7,6498)
Estadístico F	22,64	27,79*	47,80*
			151,33*
			105,46*

Notas: \* estadísticamente significativo al 1% de confianza; \*\* estadísticamente significativo al 5% de confianza; † estadísticamente significativo al 10% de confianza.

En cada modelo se utilizan 7 variables dicótomas para controlar los efectos específicos del semestre en que se hizo la evaluación; el grupo de referencia es la evaluación del último semestre. Los errores estándar aparecen entre paréntesis. En la lista de "Instrumentos" se incluyen todas las variables independientes, el intercepto y los indicadores de calidad del estudiante: ÍAG 2, ÍAG 3, ÍAG 4 e ÍAG 5. Las variables cardinales (edad del profesor, número de estudiantes presentes en la evaluación y la calificación esperada en el curso) están estandarizadas.

## EL MODELO ESTADÍSTICO

Para estimar los efectos del rango, el grado académico, la hora y el semestre en que se realizó la evaluación, así como del género del profesor y su relación con las variables anteriores se utiliza un modelo de variables binarias que equivale a un modelo de múltiples efectos fijos (Greene, 1997, 615-618). El efecto de todas las variables cualitativas se puede analizar por los desplazamientos del intercepto de la ecuación estimada. El modelo que se estima con el método de MC2E es  $SET_{ij} = \gamma + \beta^T x_{ij} + u_{ij}$ . El vector  $x_{ij}$  contiene  $k$  variables (sin incluir el intercepto) relacionadas con las características de los estudiantes, de los profesores, de los departamentos y de los cursos. Además de esas variables se incluyen las siguientes: edad del profesor, número de estudiantes que realizó la evaluación,  $\hat{C}_{ij}$  y  $\hat{C}_{ij}^2$ . Se supone que los errores  $u_{ij}$  están normalmente distribuidos.

## RESULTADOS Y DISCUSIÓN

El primer aspecto que se debe considerar es la posibilidad de que  $C_{ij}$  sea una variable endógena. La prueba de Hausman así lo confirma, pues la hipótesis nula se rechaza al más alto nivel de confianza. Por tanto, se adopta el método de MC2E (ver al Apéndice donde se incluyen las dos ecuaciones estimadas para llevar a cabo la prueba). Una vez resuelto el problema de elegir el método econométrico apropiado, se procede a modelar  $SET_{ij}$  utilizando diferentes especificaciones con diferentes conjuntos de variables independientes.

En el modelo 1 las variables independientes son el número de estudiantes presentes, la hora de la evaluación y las variables que definen las características de los profesores: grado, rango académico, género y edad (como aproximación de su experiencia docente). Excepto los coeficientes de las variables “catedrático” y “hora de la evaluación”, todos los demás son estadísticamente significativos. Comparados con un profesor con rango de “catedrático auxiliar”, los instructores y los catedráticos asociados reciben primas en sus evaluaciones esperadas de 4,65 y 3,91 puntos, respectivamente.

Los coeficientes del término lineal (positivo) y del término cuadrático (negativo) de la variable “edad” se comportan como se esperaba, es decir, las evaluaciones de los profesores muestran rendimientos marginales decrecientes con respecto a la edad (experiencia). La función es cóncava y alcanza el máximo en 40,21 años (menor que el promedio de la edad de la muestra: 45,21 años). Las profesoras

reciben una evaluación mayor que la de sus colegas varones en una prima equivalente a 5,29 puntos.

Por otro lado, un incremento de una desviación estándar en el número de estudiantes presentes en la evaluación eleva la puntuación esperada del profesor en 2,47 puntos. Este resultado es contrario a los que reporta la literatura internacional. Cabe destacar que en la UPR-B no se dictan cursos de múltiples secciones en grandes auditorios, como sí sucede en los pregrados de las universidades dedicadas principalmente a la investigación.

El modelo 2 analiza la posibilidad de que durante el período estudiado haya habido cambios estructurales que incidan en la evaluación estudiantil. Para ello se utilizan variables binarias (0, 1) que identifican el semestre y el año en que se hizo la evaluación. Su inclusión no introduce cambio alguno en la significación ni en el signo de las variables originales.

En el modelo 3 se usa el mismo artificio para clasificar los departamentos académicos y estimar el impacto de los efectos fijos específicos del departamento (cursos) en que enseña cada profesor sobre la evaluación esperada. Esto modifica el signo y la significación de algunas de las variables. Por ejemplo, la edad del profesor se vuelve insignificante, la variable género pasa de signo positivo (prima = 5,45 puntos) a negativo (penalidad = -3,01 puntos), y la hora de la evaluación se torna significativa con signo positivo: las evaluaciones realizadas en la mañana se asocian con primas de 1,45 puntos.

El departamento tiene un efecto significativo en la evaluación estudiantil. Si todo lo demás es constante, el hecho de ejercer la cátedra en un departamento está asociado con primas (coeficientes de regresión positivos) o penalidades (coeficientes negativos) en la evaluación esperada. Las penalidades varían de -41,33 puntos en Tecnologías de Ingeniería a -4,66 puntos en Educación. En Sistemas de Oficina, la evaluación esperada tiene una “prima” de 5,67 puntos. Puesto que el departamento académico en que enseña cada profesor es invariante en el tiempo y no hay razón alguna para suponer que la capacidad para transmitir conocimientos a los estudiantes varía según la disciplina, este resultado es una evidencia a favor de la hipótesis de que el proceso de evaluación estudiantil es sesgado.

Por otro lado, comparados con los “catedráticos auxiliares”, los instructores y los catedráticos asociados reciben primas de 2,90 y 3,77 puntos, respectivamente. Los coeficientes de las variables “catedrático” y “doctorado” no son significativos.

En el modelo 4 se incorpora como variable independiente la calificación esperada por el estudiante en el curso para determinar su impacto sobre la evaluación del profesor. Ahora 4 variables se vuelven estadísticamente significativas (“catedrático”, “doctorado”, “edad” y “educación”) y 9 dejan de serlo (“hora de la evaluación”, semestre “2”, “4”, “5” y “6”, “Física”, “Química”, “Biología” y “Español”). Además, cambian los signos y la magnitud de los coeficientes. Se incluye un término cuadrático de la calificación esperada para detectar posibles curvaturas en la relación. La evidencia estadística es consistente con tal conjetura pues las evaluaciones estudiantiles son convexas con respecto a la calificación que esperan los estudiantes. La evaluación del profesor tiende a aumentar a una tasa creciente a medida que aumentan las calificaciones que esperan sus estudiantes, es decir, el profesor  $j$  puede “comprar” evaluaciones cada vez mejores “alentando” expectativas de calificaciones más altas en cada estudiante  $i$ . Este resultado concuerda con los que reporta la literatura.

La edad del profesor como indicador de experiencia es estadísticamente significativa y exhibe el comportamiento esperado (cóncavo, lo que implica rendimientos marginales decrecientes). No obstante, la parábola alcanza el máximo a los 35 años, y a partir de esa edad la evaluación de los profesores es penalizada por cada año de “experiencia” docente. Por último, las profesoras tienen una desventaja con respecto a sus colegas varones pues debido al género su evaluación esperada disminuye en 6,88 puntos.

Para aislar el efecto del género sobre la evaluación del profesor se especificó y estimó el modelo 5<sup>5</sup>. Se pretende estimar los efectos sobre las evaluaciones de la interacción del género con el departamento académico en que enseña el profesor (cursos), su grado y su

<sup>5</sup> Mediante el modelo  $Y_{ij} = \dots + \alpha F + \beta \text{Curso}_h + \gamma(F \cdot \text{Curso}_h) + \dots$  se busca conocer los efectos específicos del curso  $h$  sobre la evaluación del profesor  $j$ . El efecto total cuando lo dicta un hombre, ( $F = 0$ ), se encuentra mediante la derivada parcial  $\partial Y_{ij} / \partial \text{Curso}_h |_{F=0} \approx \beta + \gamma F = \beta + \gamma(0) = \beta$ .  $\beta$  mide la tasa instantánea de cambio en  $Y_{ij}$  cuando el curso  $h$  lo dicta un profesor (efecto específico del curso  $h$  cuando es dictado por un profesor), mientras que  $(\beta + \gamma)$  mide esa misma tasa cuando lo dicta una mujer. La única excepción es el programa de Sistemas de Oficina (S. O.) pues todas las evaluaciones corresponden a mujeres ( $\Rightarrow F \cap S. O. = \emptyset$ ). Los efectos específicos del género femenino sobre la evaluación se computan así:  $\partial Y_{ij} / \partial F \approx \alpha + \gamma_{Ch} \text{Curso}_h + \gamma_r \text{Rango} + \gamma_g \text{Grado} + \dots$ . Es decir, en vez de utilizar una variable dicótoma aislada para medir la prima (penalidad) asociada al género femenino, como en el modelo 3, aquí se mide en cada uno de los programas académicos en que una mujer puede dictar cátedra, tomando en consideración su rango, grado académico, etc.

rango académico<sup>6</sup>. De los 52 coeficientes de regresión estimados, 31 son estadísticamente significativos (el 59,62%), mientras que el 57,14% de los coeficientes de interacción entre género y departamento académico es estadísticamente significativo (“Matemáticas” -12,59, “Computación” 21,22, “Electrónica” 15,88, “Administración de Empresas” 21,20, “Inglés” 12,35, “Español” 28,24, “Humanidades” 13,19, y “Educación Física” 12,47). Y 7 de ellos son positivos, lo que implica ventajas comparativas para las profesoras.

Consideremos el caso de “Matemáticas” (MATE). La razón de cambio de la evaluación es  $\partial \text{SET}_{ij} / \partial \text{MATE} = [-12,63 - 12,59*(F)]$ . Un profesor de este curso esperaría una penalidad de -12,63 puntos frente a la evaluación esperada de un profesor de Sociales (grupo de referencia). Pero si el curso es dictado por una profesora la penalidad esperada es mayor: -25,22 puntos  $[-12,63 - 12,59*(F=1) = -25,22]$ . En todos los demás casos las profesoras tienen ventajas comparativas.

En el caso de “Computación” la penalidad esperada por los varones es de -34,90 puntos y la esperada por las mujeres es de -13,68 puntos  $[-34,90 - 21,22*(F=1) = -13,68]$ . En “Administración de Empresas” (ADEM) los profesores esperan una penalidad de -26,18 puntos, mientras que la esperada por sus colegas mujeres es de sólo -4,98 puntos  $[-26,18 + 21,20*(F=1) = -4,98]$ . En el Departamento de “Inglés” los profesores esperan una penalidad de -6,69 puntos, mientras que las profesoras esperan una primas de 5,66 puntos  $[-6,69 + 12,35*(F=1) = 5,66]$ . En “Sistemas de Oficina” sólo se evaluaron mujeres y su prima esperada es de 20,81 puntos. Los datos de los demás cursos se pueden construir a partir del cuadro 3.

La evaluación del profesor está significativamente influida por su grado y rango académico. Los coeficientes de las interacciones entre las variables “género”, “rango” y “grado académico” no son significativos. La prueba de Wald no permite rechazar la hipótesis nula de que dichas interacciones son simultáneamente iguales a cero. No obstante, los efectos principales de las variables “instructor”, “catedrático asociado” y “doctorado” son estadísticamente significativos. Comparados con un “catedrático auxiliar” los “instructores” y los “catedráticos asociados” reciben primas de 4,88 y 8,68 puntos, respectivamente. Por su parte, los profesores con grado de “doctorado” son penalizados con una reducción en su evaluación esperada de 4,81 puntos (el grupo

<sup>6</sup> Para una revisión de la literatura sobre el tema, ver Centra y Gaubatz (2000).

de referencia está integrado por profesores con grado universitario o de maestría).

Si bien el coeficiente del término lineal de la variable “edad” no es significativo, una prueba de Wald lleva a rechazar la hipótesis nula de que ambos coeficientes (lineal y cuadrático) son iguales a cero. Aunque la relación tiene la curvatura esperada (cóncava), la parábola alcanza el máximo a los 26 años, lo que implica que la “experiencia” docente es penalizada todo el tiempo. Aun cuando la penalidad anual es estadísticamente significativa, su magnitud es muy pequeña.

### Cuadro 3

Cambios en la evaluación esperada inducidos por los efectos específicos de los departamentos y de sus interacciones con la variable de género ( $\partial SET_{ij} / \partial (\text{Variable})$ )

Variable	Mujer	Hombre
Ciencias de la computación	-13,68	-34,90
Electrónica	-8,11	-23,99
Administración de empresas	-4,98	-26,18
Inglés	5,66	-6,49
Español	-2,09	-30,33
Educación física	-8,98	-21,45
Sistemas de oficina	20,81	ND
Matemáticas	5,47	18,06

Sólo se incluyen las variables cuyos efectos principales, así como sus interacciones con la variable género, son estadísticamente significativos. ND: en los cursos de sistemas de oficina no se evaluó a profesor alguno.

Fuente: cálculos de los autores a partir de los coeficientes de regresión del cuadro 2.

### Cuadro 4

Primas o penalidades esperadas por las profesoras ( $\partial SET_{ij} / \partial (\text{Género})$ )

Departamentos académicos (cursos)	Primas o penalidades esperadas
Matemáticas	-25,22
Ciencias de la computación	8,59
Electrónica	3,25
Administración de empresas	8,57
Inglés	-0,28
Español	15,61
Humanidades	0,56
Educación física	-0,16
Sistemas de oficina	8,18

Fuente: cálculos de los autores a partir de los coeficientes de regresión del cuadro 2.

Se pueden utilizar procedimientos similares para aislar el efecto específico del “género” sobre la evaluación esperada cuando las demás variables permanecen constantes. El efecto total del “género” sobre la evaluación se puede descomponer en dos: “efecto principal” más “efecto de las interacciones”. El primero es negativo (-12,63 puntos),

pero en 7 de 8 casos el segundo es positivo. Lo que da a las profesoras una ventaja comparativa con respecto a sus colegas varones (ver el cuadro 4). La hora de la evaluación no es significativa, pero la evaluación esperada aumenta en 1,93 puntos por cada incremento de una desviación estándar en el número de estudiantes presentes.

La relación entre la evaluación del profesor y la calificación que esperan los estudiantes en su curso sigue siendo convexa, y es sin duda alguna la variable más importante. Por ejemplo, un incremento en la calificación esperada de “F” a “D” aumenta la evaluación esperada por el profesor en 36 puntos; de “D” a “C”, en 52 puntos; de “C” a “B”, en 69 puntos; mientras que un incremento de “B” a “A” la aumenta en 86 puntos. Así, su impacto es enorme y significativamente mayor que los reportados en estudios anteriores.

Por ejemplo, McPherson (2006), utilizando el método de MCO, reporta una relación lineal de 34 puntos por el paso de una calificación menor a la inmediatamente superior. Isely y Singh (2005) estiman una relación de 21 puntos por cada incremento entre dos calificaciones consecutivas, mientras que Krautmann y Sander (1999) lo estiman entre 34 y 56 puntos. Los resultados de este estudio confirman la conclusión de que los profesores pueden “comprar” mejores evaluaciones creando entre sus estudiantes unas expectativas de calificaciones cada vez mayores.

Puesto que los estudios que citamos se limitan al uso de MCO, sus resultados están sesgados hacia abajo (una excepción es Krautmann y Sander, 1999). Como se indicó al comienzo, la incorporación de variables endógenas entre las variables independientes lleva a que el método de MCO arroje estimadores poco confiables (sesgados e inconsistentes). En esas circunstancias, el método apropiado es el de MC2E. Como en este estudio se utilizan las técnicas econométricas apropiadas, es de esperar que los resultados sean más confiables.

## RECAPITULACIÓN

Desde sus inicios en 1927, el proceso de evaluación estudiantil de los profesores de las facultades ha sido objeto de controversias. En la literatura existen dos posiciones opuestas acerca de su adecuación y validez. Algunos investigadores sostienen que los resultados son objetivos y confiables. Mientras que sus críticos argumentan que son sesgados. En este artículo se estudia la situación imperante en la UPR-B y se hace un seguimiento longitudinal a las evaluaciones de



187 profesores de tiempo completo durante 8 semestres del período 1998-1999 a 2003-2004.

El estudio intenta determinar de qué manera los efectos específicos de las características de los profesores, los estudiantes y los cursos inciden en la evaluación esperada del profesorado. Para analizar el impacto de las variables cualitativas y de las variables cardinales cuyo recorrido es muy limitado se utilizan “variables binarias” que indican la dirección y el impacto que tiene su presencia o ausencia en la evaluación esperada. La especificación es equivalente a un modelo de efectos fijos múltiples. Las variables independientes continuas se incorporan en forma normalizada.

La evidencia estadística indica que la calificación que espera el estudiante  $i$  en el curso en que evalúa al profesor  $j$  es endógena, y se descarta el método de MCO; por ello se utiliza el método de MC2E. Después se estimaron 5 versiones diferentes del modelo, cada una de las cuales incorpora sucesivamente los conjuntos de variables independientes que definen las características de los profesores, los estudiantes y los cursos.

Los resultados revelan que en la evaluación estudiantil influyen significativamente tales características. Su presencia (o su ausencia) resulta asociada con “primas” o “penalizaciones” significativas en la evaluación esperada del profesorado. Este resultado es una evidencia a favor de la conjetura de que el proceso está viciado. Un vistazo al cuadro 3 confirma que en todos los casos las profesoras tienen ventajas comparativas con respecto a sus colegas varones.

Se detectan sesgos semejantes cuando se analizan los cambios esperados en la evaluación por razones de género ( $\partial Y_{ji} / \partial \text{Género}$ ). Los resultados se reportan en el cuadro 4. Si todo lo demás es constante, las profesoras tienen ventajas comparativas en 6 de los 9 cursos que examinamos.

Por último, existe una relación convexa y significativa entre la calificación que esperan los estudiantes y la evaluación de sus profesores. Por ejemplo, un incremento de “F” a “D” aumenta la evaluación esperada del profesor en 36 puntos; mientras que si el incremento es de “B” a “A” el incremento esperado es de 86 puntos. Esta relación es estadísticamente robusta y concuerda con la conjetura del “shopping-around” que se discute en la literatura internacional, es decir, que los profesores pueden “comprar” mejores evaluaciones estudiantiles si despiertan expectativas de calificaciones crecientes entre sus estudiantes. Por lo tanto, la posibilidad de que los estudiantes y los profesores se sientan motivados a involucrarse en algún tipo de “negociación” que

les dé un beneficio mutuo es latente. Sus efectos potenciales sobre problemas que se discuten ampliamente en otros países, como el de inflación de las calificaciones, no pueden ser ignorados en nuestro entorno universitario.

## APÉNDICE

### Prueba de Hausman para verificar si $C_{ij}$ es endógena

Variables	Ecuación de regresión para predecir la:		
	$C_{ij}$	$C_{ij}^2$	SET $_{ij}$
Constante	0,3192* (0,0415)	0,1059 (0,2276)	489,8402* (4,9103)
Edad	-0,1116** (0,0489)	-0,6264** (0,2680)	4,1047 (2,7691)
(Edad) <sup>2</sup>	0,1181* (0,0477)	0,6364** (0,2614)	-7,4922* (2,7007)
Estudiantes presentes	-0,0110† (0,0067)	-0,0759** (0,0366)	1,9250* (0,3782)
ÍAA 2	0,2266* (0,0275)	0,2702† (0,1508)	
ÍAA 3	0,4723* (0,0221)	1,1506* (0,1208)	
ÍAA 4	0,6622* (0,0248)	2,3172* (0,1357)	
ÍAA 5	0,9476* (0,0236)	4,1289* (0,1292)	
Calificación esperada $C_{ij}$			94,7010* (1,6827)
$(C_{ij})^2$			13,8491* (1,4456)
Residuos de la ecuación: $C_{ij}$			-51,9281* (4,7548)
Residuos de la ecuación: $C_{ij}^2$			-3,5545* (0,9108)
R <sup>2</sup>	0,71	0,58	0,26
Estadístico F	796,50*	457,50	112,69

Notas: \* estadísticamente significativo al 1% de confianza; \*\* estadísticamente significativo al 5% de confianza; † estadísticamente significativo al 10% de confianza.

En todos los modelos se utilizan variables dicótomas para controlar los efectos específicos de los programas académicos y rangos de los profesores, de la hora y semestre de la evaluación, así como del género del profesor y sus interacciones con los programas académicos y los rangos.

## REFERENCIAS BIBLIOGRÁFICAS

1. Algozzine, B. et al. "Student Evaluation of College Teaching", *College Teaching* 52, 2004, pp. 134-141.
2. Anderson, K. H. y J. J. Siegfried. "Gender Differences in Rating the Teaching of Economics", *Eastern Economic Journal* 23, 1997, pp. 347-357.
3. Becker, W. E. y M. Watts. "How Departments of Economics Evaluate Teaching", *American Economic Review* 89, 1999, pp. 344-349.

4. Bosshardt, W. y M. Watts. "Comparing Student and Instructor Evaluations of Teaching", *Journal of Economic Education* 32, 2001, pp. 3-17.
5. Centra A. J. y N. B. Gaubatz. "Is There Gender Bias in Student Evaluations of Teaching?", *The Journal of Higher Education* 70, 2000, pp. 17-33.
6. Cohen, P. A. "An Updated and Expanded Meta-analysis of Multisection Validity Studies", paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1986.
7. Emery, C. R.; T. R. Kramer y R. G. Tian. "Return to Academic Standards: A Critique of Student Evaluations of Teaching Effectiveness", *Quality Assurance in Education* 11, 2003, pp. 37-46.
8. Finegan, T. A. y J. J. Siegfried. "Are Students Rating of Teaching Effectiveness Influenced by Instructors' English Language Proficiency?", *American Economist* 44, 2000, pp. 17-29.
9. Germain, M-L. y T. A. Scandura. "Grade Inflation and Student Individual Differences as Systematic Bias in Faculty Evaluations", *Journal of Instructional Psychology* 32, 2005, pp. 58-67.
10. Greene, W. H. *Econometric Analysis*, New York, Prentice Hall, 1997.
11. Isely, P. y H. Singh. "Do Higher Grades Leads to Favorable Student Evaluations?", *Journal of Economic Education* 36, 2005, pp. 29-42.
12. Krautmann, A. C. y W. Sander. "Grades and Student Evaluations of Teachers", *Economics of Education Review* 18, 1999, pp. 59-63.
13. Hamilton, L. C. "Grades, Class Size, and Faculty Status Predict Teaching Evaluations", *Teaching Sociology* 8, 1980, pp. 47-62.
14. Hausman, J. A. "Specification Tests in Econometrics", *Econometrica* 46, 1978, pp. 1251-1272.
15. Johnson, V. E. *Grade Inflation: A Crisis in College Education*, New York, Springer-Verlag, 2003.
16. Marsh, H. W. "Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research", *International Journal of Educational Research* 11, 1987, pp. 253-388.
17. Matos-Díaz, H. "Sobre la posibilidad de inflación de notas en la UPR-Bayamón", *Revista de Ciencias Sociales* 15, 2006, pp. 6-29.
18. McPherson, M. A. "Determinants of How Students Evaluate Teachers", *Journal of Economic Education* 37, 2006, pp. 3-20.
19. Moore, T. "Teacher Evaluations and Grades: Additional Evidence", *Journal of American Academy of Business* 9, 2006, pp. 58-62.
20. Neger, J. L.; W. Viney y G. Riedel. "Student Rating of Teaching Effectiveness: Use and Misuse", *The Midwest Quarterly* 38, 1997, pp. 218-233.
21. Remmers, H. H. "The Purdue Rating Scale for Instructors", *Educational Administration and Supervision* 6, 1927, pp. 399-406.
22. Sailor, P.; B. R. Worthen y E-H. Shin. "Class Level as a Possible Mediator of the Relationship between Grades and Student Ratings of Teaching", *Assessment and Evaluation in Higher Education* 22, 1997, pp. 261-269.

23. Simpson, P. M. y J. Siguaw. "Student Evaluations of Teaching: An Exploratory Study of the Faculty Response", *Journal of Marketing Education* 22, 2000, pp. 199-213.
24. Stringer, M. y P. Irwing. "Students' Evaluations of Teaching Effectiveness: A Structural Modeling Approach", *British Journal of Educational Psychology* 68, 1998, pp. 409-426.
25. Wachtel, H. K. "Student Evaluation of College Teaching Effectiveness: A Brief Review", *Assessment and Evaluation in Higher Education* 23, 1998, pp. 191-211.