

## LA VALORACIÓN DE LA ACTIVIDAD DOCENTE Algo más que la opinión de los estudiantes

*F. Sánchez, R. Rubio, E. Alonso y K. Retamal*

Flor Sánchez Fernández es Vicerrectora de Planificación y Calidad en la Universidad Autónoma de Madrid. Rosa Rubio Llorente es Técnico en Evaluación en el Gabinete de Estudios y Evaluación Institucional. Ester Alonso Velasco es Directora del Gabinete de Estudios y Evaluación Institucional. Karina Retamal Soto es Becaria de Investigación en el Gabinete de Estudios y Evaluación Institucional. Todas ellas en la Universidad Autónoma de Madrid.

### **Introducción**

La valoración adecuada de la docencia es una reivindicación sostenida en gran parte de la comunidad universitaria que reclama que la labor docente consiga un reconocimiento equiparable al que ha obtenido la labor investigadora, a través de los sexenios de investigación. Una vez aceptado

el desafío, no está, sin embargo, bien definido aún el proceso para medir la calidad de la actividad docente.

En los últimos años, en el ámbito universitario, quizá como complemento a todos los cambios ocasionados por la incorporación al Espacio Europeo de Educación Superior, la LOU y su modificación, la LOMLOU (2008), se ha acentuado la necesidad de desarrollar procedimientos para la evaluación integral de la actividad docente, y de definir procedimientos que contemplen de manera global e integradora todas las tareas que forman parte de la labor docente, que va más allá de la impartición de las clases, y la diversidad de agentes que pueden aportar información al proceso de evaluación, que son más que los estudiantes y las encuestas de opinión que de forma voluntaria algunos de ellos cumplimentan.

El Real Decreto 1312/2007, de 5 de octubre, por el que se establece la acreditación nacional para el acceso a los cuerpos docentes universitarios, señala en el apartado 2B de su Anexo, como uno de los méritos

que valorar por la comisión creada a tal efecto, la calidad de la actividad docente. Así mismo, en el capítulo 7 (apartado 2B) de la "Guía de Ayuda" del programa ACADEMIA, programa de acreditación del profesorado para acceder a las distintas categorías docentes propuestas en la LOU, elaborado por la ANECA, se señala que se le pedirá a los solicitantes que incluyan dos certificados que tendrán que emitir las universidades: un certificado de la dedicación docente y un certificado de la calidad de la actividad docente.

Aunque formalmente la calidad de la docencia está incluida entre las propuestas estratégicas de las universidades, las promociones se alcanzan por los resultados de la investigación, tanto en España como en el espacio universitario europeo. Pero las universidades se están viendo obligadas a poner en marcha procesos de evaluación más amplios, en la medida que la evaluación de la calidad de la actividad docente pasa a ser considerada un indicador de calidad, incluso con peso en los modelos de financiación pactados con algunas comunidades autónomas.

Un principio fundamental para diseñar e implementar procesos de evaluación coherentes requiere definir los fines de los mismos. Y en este sentido cabe decir que, en términos generales, la evaluación de la actividad docente no ha sido tenida en cuenta debidamente en los procesos de promoción para acceder a las distintas figuras de la carrera académica. Normalmente, no es suficiente para acceder a una cátedra tener buenas valoraciones docentes de los estudiantes, pero para algunos parece ser suficiente tener un buen curriculum de investigación. El profesorado universitario asume que su promoción u otro reconocimiento depende más de su labor investigadora, "fácilmente medible", que de su calidad docente, información inespecífica, rumoroso, que se deriva de las opiniones de los estudiantes (MacAlpine, 2001).

#### *Las encuestas de los estudiantes como instrumento de evaluación*

El procedimiento de evaluación de la actividad docente más extendido en la universidad española corresponde a la evaluación realizada voluntariamente por los estudiantes, cuando opinan sobre la labor docente de los profesores. La evaluación del estudiante se ha convertido en un proceso ampliamente aceptado como forma de evaluar a los profesores en el contexto universitario nacional e internacional (Stapleton y Murkison, 2001).

Los resultados de la investigación sobre fiabilidad y validez de las encuestas de opinión de los estudiantes son dispares, algunos de ellos mencionados a continuación. Por ejemplo, West (1988) encuentra que los estudiantes fueron solo moderadamente consistentes en sus evaluaciones de los profesores en distintos momentos temporales. Son muchos los factores que podrían afectar a las evaluaciones: el tamaño de la clase (Min y Baozhi, 1998), la hora del día (Greenwald y Gilmore, 1997), la raza y el atractivo del profesor (Rao, 1995) y las evaluaciones espera-

das (Marsh y Roche, 2000), la reputación previa del profesor (Griffin, 2001), el nivel de exigencia de la asignatura, ya sean las evaluaciones presenciales o a través de la red (Felton, Mitchell y Stinson, 2004). Cabe destacar que las calificaciones que reciben los estudiantes por parte del profesor evaluado es uno de los aspectos más controvertidos. Para algunos, el grado de indulgencia del profesor, tal como es percibido por los estudiantes, se traduce en más altas evaluaciones por parte de los estudiantes (Griffin, 2004; Isely y Singh, 2005). La discrepancia o diferencia entre las calificaciones obtenidas y las esperadas influye negativamente sobre las evaluaciones del profesor, lo que coincide con la hipótesis del *self-serving* de Gigliotti y Buchtel (1990), según la cual los estudiantes tienden a penalizar con bajas puntuaciones a los profesores por este motivo, lo que podría llevar a la inflación de las calificaciones (Greenwald, 1997; Pfeffer y Fong, 2002), aunque también puede ocurrir lo contrario, que se valore mal a profesores que tienen bajos niveles de exigencia o que imparten materias muy fáciles (Marsh, 1980; Marsh y Roche, 1997). Barth (2008), apunta que en general hay un impacto negativo sobre la evaluación general de profesor en aquellos profesores que ponen estándares más rigurosos, pero que los profesores pueden corregir este pequeño efecto negativo si proporcionan ayuda y tutoría a los estudiantes. El entusiasmo del profesor por trabajar con los estudiantes fuera de la clase puede compensar este nivel de exigencia o dificultad de la asignatura. Moore (2006), apunta a una mayoría de estudios que constatan que las evaluaciones son difíciles de modificar asignando altas calificaciones al trabajo de los estudiantes o reduciendo la exigencia de los programas formativos.

Por otra parte, y de forma complementaria, un número importante de estudios dan soporte a la fiabilidad y validez de las encuestas como medidas de la actuación docente y muestran empíricamente que las evaluaciones de la docencia por parte de los estudiantes son fiables y robustos indicadores de la efectividad de la labor docente. Algunos encuentran coincidencia entre las valoraciones de los estudiantes y las que hacían los propios profesores, por ejemplo, Gamliel y Davidovitz (2005) o Drew, Burroughs y Nokovish (1987). Kwan, (1999), apoya que la opinión de los estudiantes es un indicador razonable de la calidad de la docencia hasta un 70%, el 30% restante estaría influido por los sesgos introducidos por otros factores como el tamaño de la clase, o el material del curso.

En síntesis, considerando conjuntamente el tema de los estudiantes y el tema de las encuestas conjuntamente, las opiniones se dividen entre quienes señalan que las encuestas tienen sesgos y que los estudiantes son incapaces de evaluar efectivamente la docencia (Wachtel, 1998), hasta los que proponen que los estudiantes son verdaderamente capaces de evaluar la actuación docente (Miller, 1998) y que una en-

cuesta correctamente aplicada es una de las mejores medidas de la efectividad de la docencia (Hobson y Talbot, 2001).

A la discusión sobre la fiabilidad y la validez de las encuestas se añade el uso que se ha hecho de los datos que aportan. Las consideraciones se han centrado casi exclusivamente en la puntuación obtenida en ítems sobre la satisfacción general con la labor realizada por el profesor, incluso cuando la encuesta incluye distintas dimensiones de planificación, coordinación o motivación. Los informes agregados a diferentes grupos de interés y las posibles clasificaciones se refieren a los ítems de satisfacción general infravalorando otros aspectos que podrían ser aprovechados para mejorar la labor docente.

En cualquier caso, y habiendo considerado los aspectos favorables y desfavorables, la valoración de la labor docente requiere algo más que la opinión de los estudiantes. Sin embargo, en la literatura revisada, amplia y pormenorizada, apenas hemos encontrado estudios que hayan incorporado otros elementos de valoración al proceso de evaluación docente, cabe citar el realizado por MacAlpine (2001) que compara la valoración de los estudiantes con la que hacen pares tras una observación en el aula y encuentra valoraciones coincidentes. Es tiempo de generar nuevos procesos para valorar la actividad docente, nuevos procedimientos que incorporen mecanismos complementarios para recoger información sobre las tareas docentes y que contemplen el reconocimiento y la valoración de la labor realizada. Es necesario incorporar al proceso de evaluación de la actividad docente a los diferentes agentes que participan en la misma. En primer lugar a los estudiantes, por supuesto, pero también al profesor que realiza la labor docente, que debe tener la oportunidad de autoevaluar su labor y los resultados que obtiene, y a los responsables académicos de la docencia, al menos.

Esta idea de una evaluación de labor docente contemplando aspectos diferentes y complementarios es el fundamento del Programa de Apoyo a la Evaluación de la Actividad Docente (DOCENTIA) promovido por ANECA (Agencia Nacional de Evaluación de la Calidad y Acreditación) para acreditar procedimientos orientados hacia la evaluación integral de la calidad docente y es el que se ha empezado a desarrollar en las universidades españolas.

En el caso de la Universidad Autónoma de Madrid, el Consejo de Gobierno aprobó poner en marcha un programa piloto para la "Identificación y Valoración de las Prácticas Docentes" que tiene como finalidad conocer mejor cómo se desarrolla la actividad docente, a la vez que identificar aquellas prácticas que puedan ser consideradas de calidad y que aporten un valor añadido a la docencia. El resultado esperado es que la institución en su conjunto tenga la posibilidad de aprender, compartiendo las prácticas docentes de calidad identificadas. El modelo de valoración de la actividad docente que hemos seguido, toma como referencia el modelo DOCENTIA de ANECA y lo adapta a las características

**Cuadro 1**  
Dimensiones y Criterios del modelo con las puntuaciones asociadas

DIMENSION	CRITERIO	PUNTUACIÓN POR FUENTES DE INFORMACIÓN				TOTAL POR CRITERIOS	TOTAL POR DIMENSIÓN
		EE	IA	IRA	BdD		
1. Encargo Docente	1.1. Dedicación docente		6			6	10
	1.2. Diversidad del encargo docente		3			3	
	1.3. Condiciones en las que se prestó el encargo docente.		-			0	
	1.4. Cumplimiento del encargo docente			1		1	
2. Desempeño Docente	2.1. Coordinación y planificación	0,5	8		1,5	10	60
	2.2. Desarrollo de la docencia	9	8	1	2	20	
	2.3. Interacción con los estudiantes	2	8			10	
	2.4. Evaluación del proceso de enseñanza-aprendizaje	3	17			20	
	2.5. Resultados de los estudiantes					0	
3. Formación, innovación, investigación docente y actividades institucionales de mejora de la docencia	3.1. Formación, innovación, investigación docente		9	1		10	20
	3.2. Actividades institucionales de mejora de la docencia		9	1		10	
4. Desarrollo de materiales didácticos	4.1. La implicación del profesor en la generación de materiales para la docencia.		9	1		10	10
<b>Puntuación total</b>		<b>14,5</b>	<b>77</b>	<b>5</b>	<b>3,5</b>	<b>100</b>	<b>100</b>

Puntuaciones directas por procedencia de los datos: Encuestas de actuación docente de los 3 últimos cursos (EE); Informe de Autovaloración del profesor (IA); Informe del responsable académico (IRA); Programas de las asignaturas impartidas en el último curso (PR) y Página Web de docencia (W)

de la propia universidad. La estructura general del modelo y las valoraciones de cada dimensión y criterio aparece en el Cuadro 1, que también muestra la distribución de puntuaciones por cada una de las fuentes de información en las dimensiones y criterios que conforman el modelo.

La propuesta de valoración considera 4 dimensiones de análisis que se concretan en 12 criterios y 70 indicadores, de los cuales 53 son puntuables y 17 aportan información de contexto. La valoración se fundamenta en la recogida y análisis de información procedente de distintas fuentes: los estudiantes, a través de la encuesta de opinión (EE), el profesor, a través del informe de autovaloración (IA); el director del departamento a partir del informe del responsable académico (IRA) y otros datos del profesor recogidos en las bases de datos institucionales. El profesor aporta información relevante para la valoración de las 4 dimensiones de análisis: dimensión 1, Encargo docente (información que posteriormente se contrasta con la recogida en las bases de datos institucionales); dimensión 2, Desempeño docente; dimensión 3, Formación, innovación y actividades institucionales de mejora de la docencia y dimensión 4, Diseño material didáctico. De igual manera, la información aportada por el responsable académico está referida a las 4 dimensiones de valoración. Los datos procedentes de los resultados de la encuesta de estudiantes se consideran únicamente en la valoración de los distintos criterios de la dimensión 2, Desempeño docente. La información recogida relativa a las dimensiones 1 y 2 (para todas las fuentes consultadas) está referida a los tres cursos previos al de la convocatoria (2003-2004; 2004-2005 y 2005-2006). La referida a las dimensiones 3 y 4 abarca un rango mayor (los seis cursos anteriores al de la convocatoria) puesto que las actividades formativas y de innovación así como la generación de materiales para la docencia no tienen el mismo carácter "acumulativo" que puede tener el desempeño docente.

La información aportada es ordenada y clasificada por los técnicos de evaluación de acuerdo a los criterios e indicadores previamente definidos en el modelo y, en una segunda fase, es valorada de forma global por un comité de expertos.

El proyecto tiene como objetivo final proponer un modelo para reconocer y valorar la actividad docente del profesorado. Por ello se aplican distintos métodos para documentar buenas prácticas docentes que puedan ser compartidas por la comunidad universitaria, así como identificar las estrategias organizadas y eficaces de actuación docente.

## **Método**

Los datos que aparecen en este trabajo corresponden a un muestra de 82 profesores que representaban las distintas categorías académicas

excepto catedráticos, siendo el 50% de los participantes mujeres y que se incorporaron de forma voluntaria en el proyecto piloto para la "Identificación y Valoración de Prácticas Docentes". La documentación relacionada con el proyecto está en la página web <http://www.uam.es/calidad/gabinete/practicasdcentes/index.htm>

### *Instrumentos y procedimiento*

Para poner a prueba el modelo de evaluación propuesto se codificó y analizó la información que aportaron los profesores en torno a la actividad docente desarrollada entre los tres y los 6 años previos al estudio, dependiendo del tipo de información requerida, así como la que aportaron los responsables académicos y los estudiantes cumplimentando los instrumentos que se describen a continuación.

*Informe de autovaloración del profesor (IA)*, mediante el que se recogió información sobre las 4 dimensiones valorables en el modelo propuesto y específicamente sobre los aspectos que se describen a continuación. Ello se hizo a través de escalas tipo Likert de 6 puntos o preguntas dicotómicas, preguntas abiertas, según la información a recopilar. La valoración concreta de cada apartado aparece en el Cuadro 1.

- 1.- *Encargo docente*: asignaturas impartidas; nº de grupos; nº total de créditos impartidos; elección de la asignatura y relación con sus líneas de investigación; condiciones en que se prestó el encargo docente y condiciones que habían influido en el desarrollo de la docencia.
- 2.- *Desempeño docente*: información referida a coordinación y planificación, a *desarrollo de la docencia*, *interacción con los estudiantes* y *evaluación* del proceso de enseñanza-aprendizaje.
- 3.- Información sobre *actividades formativas y de innovación*, así como actividades institucionales de mejora de la docencia realizadas en el periodo considerado
- 4.- Información sobre *material didáctico* desarrollado en los 6 años previos, tanto el material con ISBN o ISSN como el material desarrollado para la docencia que no había sido publicado.
- 5.- Información sobre acciones realizadas y que el profesor consideraba destacables y que habían supuesto una mejora en su docencia.

*Informe del responsable académico (IRA)*. Se solicitó a los directores de los departamentos a los que pertenecían los profesores participantes un informe personalizado de cada uno de éstos, utilizando un instrumento diseñado específicamente que planteaba cuestiones sobre los temas que se mencionan a continuación. Para responder se utilizaron escalas tipo Likert de 6 puntos, o preguntas dicotómicas.

- 1.- Información sobre el encargo docente del profesor a través de 8 cuestiones.
- 2.- Información sobre el desempeño docente
- 3.- Información sobre participación del profesor en actividades formativas de innovación docente y en actividades institucionales de mejora de la docencia,
- 4.- Información sobre desarrollo de materiales didácticos por parte del profesor.
- 5.- Valoración global de la actividad realizada por el profesor.
- 6.- Otras observaciones que el responsable académico considerara oportuno realizar.

*Encuesta de opinión de los estudiantes (EE)* sobre la actuación docente de los profesores participantes en los 3 últimos años. La encuesta tiene carácter institucional y se pasa para cada profesor al finalizar la docencia en cada una de las asignaturas impartidas y en cada grupo. La encuesta incluye 25 preguntas que abordan aspectos sobre la organización docente, la planificación de la docencia, la interacción con los estudiantes y un apartado de satisfacción general con la labor docente.

El análisis de la información se llevó a cabo en dos fases. En la primera, realizada por técnicos de evaluación del Gabinete de Estudios, se valoraron los indicadores asociados a información procedente de los informes del profesor, el responsable académico, las encuestas de los estudiantes, etc. En la segunda fase, la información aportada por los participantes fue valorada por un comité mixto de expertos, que valoró el contenido de las preguntas abiertas del informe de autovaloración y la documentación enviada por los participantes, tras un periodo de formación previo. El comité de expertos estuvo integrado por profesores procedentes de la propia institución y por expertos externos a propuesta de la Agencia de Calidad, Acreditación y Prospectiva de la Comunidad de Madrid. Finalmente el comité quedó constituido por 12 integrantes: 8 pertenecientes a la UAM y 4 pertenecientes a otras universidades públicas madrileñas.

El análisis de esta información diversa permitió construir la Puntuación Total (PT) que es la suma de las puntuaciones alcanzadas por cada profesor para cada uno de los indicadores en cada uno de los criterios y dimensiones del modelo propuesto. Así mismo se definió una puntuación para cada una de las dimensiones y para cada uno de los criterios que contemplaba el modelo de evaluación propuesto.

El desarrollo del proyecto en su conjunto, desde la apertura de la convocatoria para que los profesores pudieran participar hasta la entrega del informe al profesor requirió casi un año de trabajo.



## Resultados

Una vez aplicado el modelo propuesto para la valoración de la actividad docente, se procedió a analizar el sentido y coherencia de la información obtenida para contrastar en qué medida se puede sostener empíricamente la propuesta de evaluación integrada de actividad docente.

Como paso previo se llevó a cabo un análisis sobre la distribución de la muestra en relación con la variable Puntuación Total (PT), que mostró una distribución normal ( $\chi^2= 0.976$ ;  $p=1,000$ ), con una media de 52,55 y una desviación típica de 13,07. Se analizó la fiabilidad global de los instrumentos utilizados. El valor Alfa de Cronbach fue de 0,70 para el total de 47 indicadores y dado el carácter experimental del proyecto, se dio por válido para continuar en la exploración, ajuste y mejora del modelo.

Teniendo en cuenta la información procedente de las distintas fuentes (informe de autovaloración del profesor (IA), informe del responsable académico (IRA) y la encuesta de los estudiantes (EE), así como las valoraciones hechas por los técnicos (puntuación directa) y las realizadas por el comité experto estructuramos los análisis siguientes.

En primer lugar, se exploró el comportamiento del modelo por dimensiones teóricas y criterios de valoración. En un segundo momento se analizó la aportación de cada una de las fuentes de información consultadas (profesor; responsable académico; estudiantes) a la puntuación total obtenida por los profesores. Posteriormente se realizó una comparación entre las valoraciones realizadas por los técnicos y las valoraciones del comité experto. Finalmente se desarrolló un análisis de contenido de la información que se aportó sobre las buenas prácticas que los profesores creían llevar a cabo con el fin de obtener los mejores resultados de su labor docente.

### *Análisis del modelo por dimensiones teóricas y criterios de valoración*

Los resultados del análisis de las puntuaciones de los participantes en el estudio para cada una de las dimensiones de análisis, aportan los datos de la Tabla 1.

**Tabla 1**  
Dimensiones de valoración de la actividad docente

	Media	SD	Mínimo	Máx.	Cuartil 1	Cuartil 2	Cuartil 3
D1: Encargo Docente (10 puntos)	6,90	1,63	1,00	10,00	6,00	6,88	7,81
D2: Desempeño Docente (60 puntos)	34,88	8,32	13,94	54,43	29,57	35,38	41,21
D3: Formación, innovación y actividades institucionales (20 puntos)	6,40	5,15	0,00	20,00	2,00	5,25	9,94
D4: Diseño material didáctico (10 pts)	4,36	2,84	0,00	10,00	1,94	4,00	5,56

Como puede observarse en la Tabla 1, las puntuaciones medias de las dimensiones 1 y 2 se sitúan en valores superiores a la media de la puntuación aportada por esas dimensiones (media dimensión 1 = 6,9 sobre un total de 10 puntos; media dimensión 2 = 34,88 sobre un total de 60 puntos). No sucede así en el caso de las dimensiones 3 y 4; cuyas medias se sitúan, en ambos casos, en valores inferiores a la mitad de la puntuación aportada por cada una de las dos dimensiones (media dimensión 3 = 6,40 sobre un total de 20 puntos; media dimensión 4 = 4,36 sobre un total de 10 puntos).

Una explicación, al menos parcial, de estos datos tendrían que ver con el carácter novedoso y retrospectivo de este proyecto. Por un lado, se valora información que no se ha venido considerando directamente vinculada con la docencia de calidad (como es el caso de participación en actividades institucionales relacionadas con la docencia) o actividades que no tienen una amplia participación, o no han tenido una amplia oferta, como es el caso de los programas de formación docente, que justifica, al menos en parte, que los profesores acumulen en sus currículum pocas acciones formativas. Por otro lado, la producción bibliográfica de los profesores suele ir encaminada a la difusión de sus resultados de investigación, que es lo que se ha venido valorando en el currículum académico, y no tanto al desarrollo de material de apoyo para la docencia.

A continuación se llevó a cabo un análisis de correlaciones para determinar las relaciones existentes entre la puntuación total obtenida por cada participante, y la puntuación alcanzada por cada uno de ellos en las distintas dimensiones del modelo: como era esperable la puntuación total (PT) correlaciona de forma significativa y positiva con las puntuaciones de las distintas dimensiones ( $r^2_{PT-D1} = 0,513$ ,  $p=0,000$ ;  $r^2_{PT-D2} = 0,884$ ,  $p=0,000$ ;  $r^2_{PT-D3} = 0,747$ ,  $p=0,000$ ;  $r^2_{PT-D4} = 0,367$ ,  $p=0,001$ ). Igualmente, las dimensiones D1; D2 y D3 correlacionan entre sí, si bien ésta relación, aunque positiva, no es muy elevada ( $r^2_{D1-D2} = 0,354$ ,  $p=0,001$ ;  $r^2_{D1-D3} = 0,361$ ,  $p=0,001$ ;  $r^2_{D2-D3} = 0,440$ ,  $p=0,000$ ). No hay correlación entre la dimensión 4 y el resto de dimensiones del modelo. Estos datos soportarían la interrelación de las actividades que constituyen la labor docente de un profesor universitario.

Para tener una mayor información sobre las relaciones entre las diversas actividades que los profesores informan realizar, se analizaron los datos a nivel de criterios de valoración, con los resultados que muestra la Tabla 2. Entre paréntesis y al lado de cada criterio su valoración en el modelo.

**Tabla 2**  
Criterios de valoración de la actividad docente

	Media	SD.	Mínimo	Máx.	Cuartil 1	Cuartil 2	Cuartil 3
Crt1.1: Dedicación Docente (6)	4,23	1,25	0,00	6,00	3,50	4,50	4,75
Crt1.2: Diversidad encargo docente(3)	1,68	0,75	0,00	3,00	1,00	1,88	2,06
Crt1.4: Cumplimiento encargo docente (1)	0,98	0,08	0,50	1,00	1,00	1,00	1,00
Crt2.1: Coordinación y planificación (10)	7,60	1,51	3,66	10,00	6,50	7,84	8,75
Crt2.2: Desarrollo docencia (20)	9,73	3,91	2,00	18,63	6,61	8,82	13,00
Crt2.3: Interacción con estudiantes (10)	5,63	2,09	1,50	9,00	4,00	6,00	7,00
Crt2.4: Evaluación (20)	11,93	3,33	2,75	19,00	10,00	12,00	14,50
Crt3.1: Formación e innovación (10)	3,45	3,20	0,00	10,00	0,75	2,25	5,63
Crt3.2: Actividade institucionales (10)	2,96	2,97	0,00	10,00	1,00	1,00	5,00
Crt4.1: Diseño material didáctico(10)	4,36	2,84	0,00	10,00	1,94	4,00	5,56

Como muestra la Tabla 2, la mayor parte de los criterios propuestos se comportan de modo similar a las dimensiones de las que forman parte. Un aspecto destacable es que el criterio Crt1.4, cumplimiento del encargo docente, prácticamente no tiene variabilidad (SD=0,08). Este criterio, recordemos que recoge la respuesta del responsable académico en cuanto a la existencia de evidencias positivas, negativas o la falta de evidencias sobre el cumplimiento de una serie de obligaciones formales por parte del profesor, siendo estas respuestas en su mayor parte muy positivas. Esto es, los directores de departamento como responsables académicos valoran, en términos generales, muy positivamente la labor docente de los profesores.

Además, se analizó la relación existente entre los criterios de valoración y la puntuación total. Se detectó una relación muy baja entre el criterio Crt1.2 y la Puntuación total (PT) ( $r^2_{PT-Crt1.2} = 0,261$ ,  $p=0,018$ ) y una ausencia de correlación significativa entre ésta última y el indicador Crt1.4.

A continuación se realizó un análisis de componentes principales con rotación "Oblimin" para determinar la contribución de cada uno de los

criterios a la puntuación total, excluyéndose del análisis aquellos criterios que tenían valores de correlación inferiores a 0,30. Se forzó el análisis a 4 componentes para explorar si el modelo empírico se ajustaba al modelo teórico. En el gráfico 1 y en la Tabla 3 aparecen los datos al respecto, incluyéndose solo los autovalores que superan el 0.5.

**Tabla 3**  
Componentes del modelo

Componente	Autovalores iniciales		
	Total	% de la varianza	% acumu- lado
1	3,078	38,481	38,481
2	1,099	13,734	52,215
3	1,005	12,563	64,778
4	0,806	10,071	74,849

En total, los 4 componentes explican el 74,8% de la varianza total. El componente 1, está conformado por tres de los criterios tradicionalmente relacionados con la actividad docente (desarrollo de la docencia, interacción con los estudiantes y evaluación de la docencia), todos ellos pertenecientes a la dimensión 2 del modelo de evaluación. El segundo componente se configura por un solo criterio (desarrollo de material didáctico) coincidente con la dimensión 4 del modelo de valoración propuesto. El tercer componente está integrado por los dos criterios pertenecientes a la dimensión 3 (actividades institucionales de mejora de la docencia y formación, innovación e investigación docente). El componente 4 está formado por dos criterios (dedicación docente y coordinación y planificación) pertenecientes a la dimensión 1 y a la dimensión 2 respectivamente. En la Tabla 4 aparecen los pesos de cada criterio en cada uno de los componentes.

**Tabla 4**  
Relaciones entre componentes y criterios

	Componente			
	1	2	3	4
Crt1.1: Dedicación Docente				0,89
Crt2.1: Coordinación y planificación				0,64
Crt2.2: Desarrollo docencia	0,53			
Crt2.3: Interacción con estudiantes	0,88			
Crt2.4: Evaluación	0,82			
Crt3.1: Formación e innovación			0,74	
Crt3.2: Actividades institucionales			0,83	
Crt4.1: Diseño material didáctico		0,97		

*Análisis de la aportación de cada una de las fuentes de información: profesor, responsable académico y estudiantes a la puntuación total (PT)*

En la Tabla 5 aparecen los resultados de estos análisis que en general muestran que la puntuación total obtenida depende simultáneamente de las tres fuentes de información.

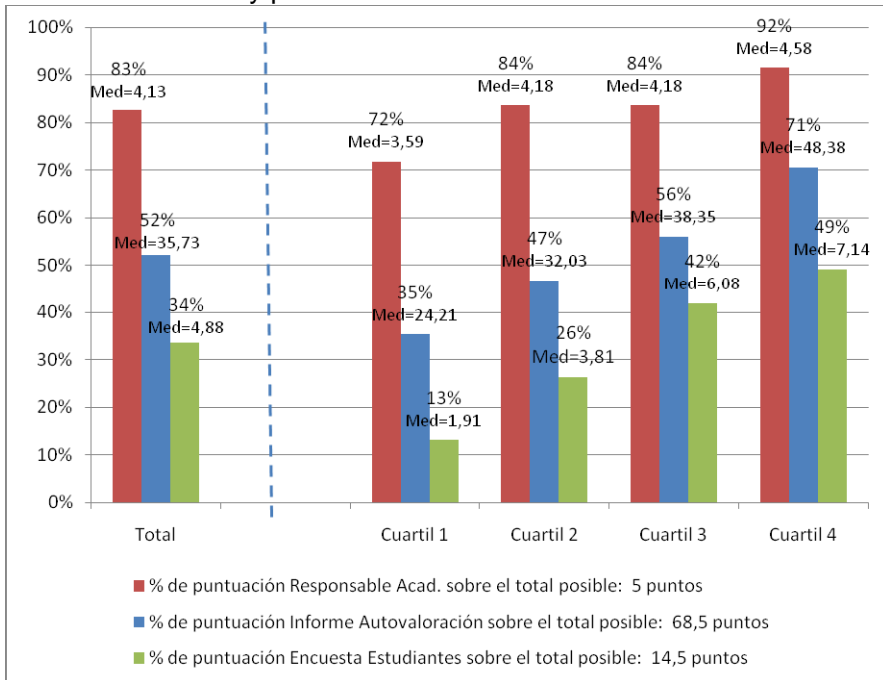
**Tabla 5**  
Aportaciones de las fuentes de información

	Media	SD	Mínimo	Máx.
Informe Autovaloración (IA) ( 68,5)	35,73	9,89	15,45	60,00
Encuesta Estudiantes (EE) (14,5)	4,88	4,06	,00	14,50
Inf.Responsable Académico IRA ( 5)	4,13	,89	,50	5,00

A partir del análisis de correlaciones de la puntuación total (PT) con las puntuaciones de las distintas fuentes, y de éstas entre sí, se aprecia, como era esperable, una relación positiva y significativa entre la puntuación total (PT) obtenida y la puntuación aportada por cada una de las fuentes consultadas ( $r^2_{PT-Puntuación IA} = 0,931$ ,  $p=0,000$ ;  $r^2_{PT-Puntuación EE} = 0,504$ ,  $p=0,000$ ;  $r^2_{PT-Puntuación RA} = 0,403$ ,  $p=0,000$ ). También correlacionan entre sí y de forma significativa, aunque en menor medida, la puntuación obtenida en el Informe de Autovaloración y la aportada por el Informe del Responsable académico ( $r^2_{Puntuación IA-Puntuación RA} = 0,332$ ,  $p=0,002$ ).

De forma complementaria, se analizó la aportación de cada una de las fuentes, en función de la posición de los participantes en la puntuación total (PT) estructurada agrupada en cuartiles. El gráfico 1 muestra el porcentaje de puntuación aportada por cada una de las fuentes en relación con el total de puntuación posible que puede aportar cada fuente.

**Gráfico 1**  
Fuentes de información y su aportación a la puntuación total y para cada uno de los cuartiles



Como era esperable, el porcentaje de puntuación aportada por el propio profesor, por el responsable académico y por los estudiantes va aumentando a través de los cuartiles. El porcentaje de valoración más bajo en relación con la puntuación máxima posible corresponde a la encuesta de estudiantes ya que ni siquiera en el cuarto cuartil el porcentaje medio alcanza el 50% de los puntos que se podrían asignar. De acuerdo con estos datos los profesores con más puntuación total son también los mejor valorados por los estudiantes, aunque tal valoración es menor que la que el profesor consigue a partir de su informe de autovaloración, lo que coincide con resultados de estudios previos.

Por el contrario, en la aportación de los responsables académicos, la media de los profesores que se encuentran en el cuarto cuartil se aproxima al 100% de los puntos posibles, e incluso los profesores que están en el cuartil 1 en relación con su puntuación total reciben algo más del 70% de la puntuación posible proveniente del informe del responsable académico. Esto querría decir que la puntuación del responsable académico no discrimina mucho ya que tanto los profesores que obtie-

nen las mejores puntuaciones como los que obtienen las peores consiguen un buen informe por parte del responsable académico.

En cuanto al informe de autovaloración de los profesores, las puntuaciones disciernen entre los profesores, no todos informan las mismas prácticas, pero también merece la pena comentar que la exhaustividad a la hora de informar en un factor que favorece positivamente la puntuación mencionada.

Para explorar si los profesores diferían en la puntuación recibida de cada una de las fuentes, en función del cuartil en que se posicionasen a partir de su puntuación total, se llevó a cabo un análisis de varianza, tomando como factor la puntuación total agrupada en cuartiles. Se comprueba la existencia de diferencias significativas entre las medias de las puntuaciones aportadas por cada una de las fuentes (puntuación del informe de autovaloración  $F=99,043$ ;  $p=0,000$ ; puntuación de encuesta de estudiantes  $F=7,547$ ;  $p=0,000$ ; puntuación responsable académico  $F=4,727$ ,  $p=0,004$ ).

Mediante la prueba Tukey, se comprobó la significación de esas diferencias intergrupo. En la tabla 6 se presentan las diferencias en aquellos cuartiles entre los que resultan significativas para cada una de las fuentes de información.

**Tabla 6**  
Diferencias significativas entre las puntuaciones medias de cada cuartil por fuente de información

Fuentes de información	Cuartil	Cuartil	Cuartil	Cuartil	
	1	2	3	4	
Informe responsable académico (IRA)	Cuartil 1			-0,99*	
	Cuartil 2				
	Cuartil 3				
	Cuartil 4	0,99*			
Informe de autovaloración (IA)	Cuartil 1		-7,82*	-14,14*	-24,12*
	Cuartil 2	7,82*		-6,31*	-16,31*
	Cuartil 3	14,14*	6,31*		-10,03
	Cuartil 4	24,12*	16,31*	10,03*	
Encuesta estudiantes (EE)	Cuartil 1		-1,90*	-4,18*	-5,23*
	Cuartil 2	1,90*			-3,33*
	Cuartil 3	4,18*			
	Cuartil 4	5,23*	3,33*		

\* La diferencia es significativa al nivel 0,05

De los resultados obtenidos podría deducirse que las 3 fuentes aportan valoraciones que van en la misma dirección, más claramente si comparamos las puntuaciones del cuartil 1 y el cuartil 4, esto es, los profesores que tienen las puntuaciones totales más bajas o más altas (cuartiles 1 y 4 respectivamente) también obtienen las puntuaciones más bajas según la aportación de cada una de las fuentes de información, especialmente para los valores procedentes de la información de estudiantes y del propio profesor. Las diferencias no son tan acusadas en las posiciones medias de la puntuación total (cuartiles 2 y 3) principalmente en lo que se refiere a las valoraciones de los responsables académicos y en menor medida a las valoraciones de los estudiantes; aunque de nuevo hay diferencias entre las medias aportada por el informe de autovaloración (información del profesor) en todos los cuartiles.

*Análisis de la comparación de las valoraciones realizadas por técnicos y comité de expertos*

Como se ha comentado la información recopilada a través de los diferentes cuestionarios y la de las bases de datos institucionales fue valorada por los técnicos de calidad, y posteriormente el comité de expertos hizo una valoración global de todos los informes y un análisis concreto de la información de carácter más cualitativo aportada por el profesor en el autoinforme de valoración. Los datos correspondientes a la comparación de estas valoraciones están en la Tabla 7.

**Tabla 7**  
Comparación de agentes de valoración (técnicos vs. comité experto)

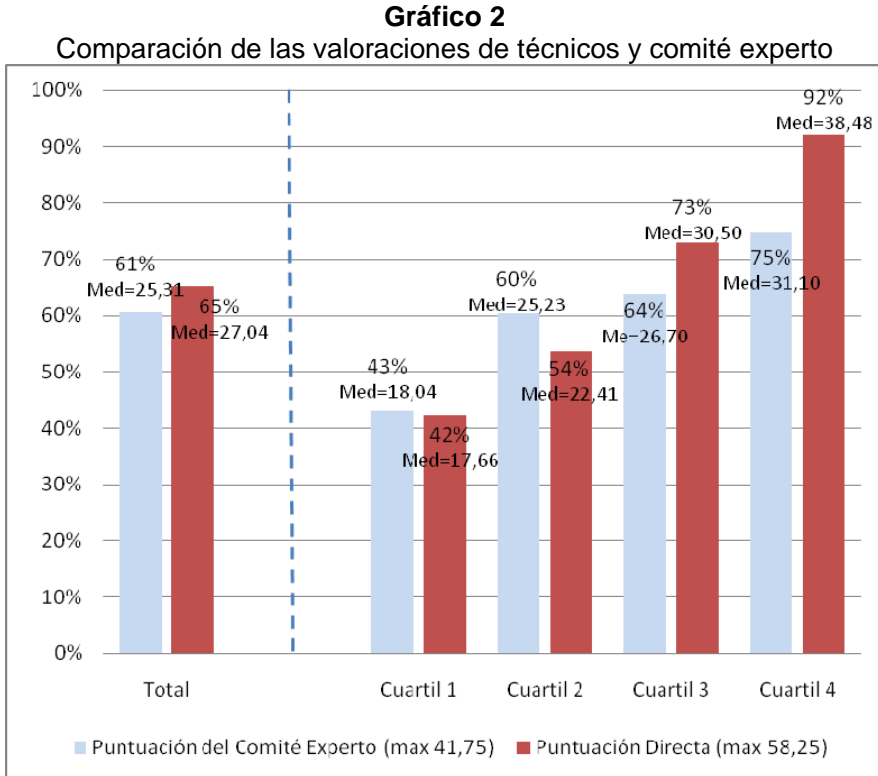
	Media	SD	Mínimo	Máximo
Valoración técnicos (máx 58,25)	27,24	9,13	10,25	49,75
Comité Experto (máx 41,75)	25,31	6,34	8,80	38,00

El análisis de correlaciones entre la puntuación total (PT) y valoración asignada por los expertos y la asignada por el comité experto muestra una correlación significativa entre cada una de estas puntuaciones y la total ( $r^2_{PT-directa} = 0,896$ ,  $p=0,002$  y  $r^2_{PT-comité-experto} = 0,769$ ,  $p=0,000$ ); y entre los dos tipos de puntuaciones entre sí ( $r^2_{comité-experto-directa} = 0,407$ ,  $p=0,000$ ).

Complementariamente, y considerando la distribución de la muestra en cuartiles según su puntuación total, comparamos la valoración de



técnicos y del comité experto en función de la posición de los participantes en la variable puntuación total. En el gráfico 2 se muestran los resultados de este análisis.



Realizamos un análisis de diferencia de medias, tomando como factor la puntuación total agrupada en cuartiles que muestra la existencia de diferencias significativas entre las medias de las puntuaciones directa ( $F=77,944$ ,  $p=0,000$ ) y la puntuación obtenida a partir de la valoración de la comisión de expertos ( $F=31,737$ ;  $p=0,000$ ).

Mediante la prueba Tukey, podemos comprobar la significación de esas diferencias intragrupo. En la Tabla 8 se representa gráficamente los cuartiles entre los que las diferencias de medias resultan significativas para cada una de los tipos de variables.

**Tabla 8**  
Diferencias significativas entre las puntuaciones medias de cada cuartil por agente de valoración

		Cuartil 1	Cuartil 2	Cuartil 3	Cuartil 4
Puntuación Comité Experto	Cuartil 1		-7,19*	-8,66*	-13,16*
	Cuartil 2	7,19*			-5,97*
	Cuartil 3	8,66*			-4,50*
	Cuartil 4	13,16*	5,97*	4,50*	
Puntuación Directa	Cuartil 1		-4,75*	-12,83*	-20,82*
	Cuartil 2	4,75*		-8,09*	-16,07*
	Cuartil 3	12,83*	8,09*		-7,98*
	Cuartil 4	20,82*	16,07*	7,98*	

\* La diferencia es significativa al nivel 0,05

De los resultados obtenidos se podría inferir que las puntuaciones procedentes de la valoración de los técnicos y de las valoraciones del comité experto tienen el mismo sentido. De modo que las puntuaciones obtenidas por los profesores que tienen las puntuaciones totales más bajas (cuartiles 1 y 2) difieren significativamente de las puntuaciones que reciben en las mismas variables los profesores que se sitúan en los valores medios y altos de la puntuación total (cuartiles 3 y 4). Esto es, los profesores que obtienen valoraciones más altas en la puntuación total son mejor valorados tanto por los técnicos como por el comité experto.

### Identificación de Buenas Prácticas docentes

Finalmente hicimos un análisis de contenido de la información aportada por los profesores en el Informe de Autovaloración, en el cual se hace una pregunta abierta final en relación con aquellas prácticas docentes que consideraban de calidad y que incluían en su labor docente. En términos generales, y considerando toda la muestra, encontramos que las mejores prácticas mencionadas tenían que ver con:

- Desarrollo de actividades dirigidas a fomentar la participación de los estudiantes en el aula y orientación práctica de la asignatura, mencionada como buena práctica por un 70,59% de los profesores que completan correctamente este apartado.
- Uso de la web y las nuevas tecnologías como herramienta para la docencia (alojamiento de material didáctico; intercambio con los

estudiantes, etc.), categoría mencionada por un 27,54% de los profesores.

- Participación en proyectos de innovación docente y/o en proyectos experimentales ECTS con resultados sobre la docencia del profesor, los cuales son asociados a la idea de buenas prácticas por el 15,69% de la muestra.
- Planificación detallada de la asignatura y coordinación con otros docentes, actividad que es mencionada como buena práctica por un 29,41% de los 51 profesores que completan correctamente este apartado:

A nivel más concreto, cuando se compara la información aportada al respecto por los profesores mejor clasificados por su puntuación total y la de los profesores con puntuaciones más bajas, no se encuentran diferencias apreciables en las prácticas informadas, lo que podría explicarse en principio por la gran diversidad que presenta la muestra en cuanto a cantidad y calidad de la información aportada. Será necesario acumular más datos al respecto para llegar a alguna conclusión sostenible.

### **Discusión y Conclusiones**

La universidad española está en la búsqueda de un modelo de valoración de la actividad docente que encaje mejor con la consideración que se quiere hacer de esta actividad. El modelo de valoración propuesto resulta coherente para valorar la actividad docente del profesor y muestra que es posible considerar otras actividades que forman parte de la labor docente de un profesor y “que se pierden” ya que no son consideradas en la aplicación de la encuesta de los estudiantes y no pertenecen al ámbito de la investigación. Los instrumentos de recogida de información muestran una fiabilidad suficiente y permiten realizar una valoración y clasificación de las actividades realizadas por los docentes.

Las puntuaciones alcanzadas por los participantes tienen un rango de variabilidad normalizado. Respecto al hecho de que las puntuaciones alcanzadas por los participantes no lleguen a la valoración máxima teórica del modelo de evaluación puede explicarse porque en un modelo basado en indicadores que se valoran de forma independiente, como es el caso, es más difícil alcanzar la puntuación máxima; puesto que se valora indicador a indicador, lo que hace que sea más frecuente la “pérdida de puntuación”, ya que si no se alcanza el indicador al 100% no recibe la puntuación máxima asignada. Tampoco es esperable que ninguno de los participantes tenga la puntuación mínima: el mero desempeño de la actividad docente pone a los profesores en disposición de realizar distintas tareas y actividades que son valoradas por el modelo.

Los datos avalan la estructura teórica en cuatro dimensiones del instrumento: “Desempeño docente”, “Encargo docente, coordinación y

planificación, "Formación y participación en Actividades Institucionales de mejora de la docencia" y "Desarrollo de material didáctico". La puesta en práctica de este tipo de proyectos puede ayudar en la recogida de información sobre las actividades realizadas por los docentes, que facilite su consideración a la hora de definir un modelo de docente.

La dimensión que explica mayor varianza es la dimensión 2, Desarrollo de la docencia, lo que se corresponde con su mayor peso en el modelo. Si bien, uno de los criterios de esta dimensión, el criterio 2.1 Coordinación y planificación en el análisis factorial realizado satura junto con el criterio 1.1 Dedicación docente. Esto puede indicar que las actividades de preparación de la asignatura (coordinación para el diseño programas y evaluación; planificación; contenidos de los documentos de los programas y distribución de los mismos) están más relacionados con el propio encargo que con el desempeño posterior (desarrollo de la docencia; interacción con los estudiantes y evaluación); por lo que una futura línea de actuación puede ser ajuste el modelo teórico recogiendo este cambio.

Las dimensiones 3 y 4 se muestran como dimensiones consistentes. La dimensión 4, Desarrollo de materiales para la docencia, se presenta como un factor independiente que podrá ir tomando mayor peso en este tipo de modelos de valoración de la actividad docente, en la medida que estos trabajos tengan un reconocimiento curricular y su producción y difusión sea mayor.

En general, las comparaciones entre las puntuaciones de las diferentes dimensiones muestran coherencia, de modo que los profesores que reciben menor valoración por los estudiantes consiguen también menor valoración en la puntuación total. Complementariamente los profesores que obtienen mejores puntuaciones obtienen también mejores valoraciones por parte de los estudiantes, dato que coincide con lo encontrado en estudios previos que sostienen que los mejores profesores son los mejor valorados por los estudiantes, incluso aunque tengan criterios de exigencia altos. Las valoraciones de los técnicos y las de los expertos son coherentes y apuntan en la misma dirección, los profesores mejor valorados, lo son tanto por los técnicos como por los expertos.

Un comentario específico requieren las valoraciones de los responsables académicos, que como hemos señalado han sido muy positivas tanto para profesores que obtienen las mejores puntuaciones como para los que obtienen las peores puntuaciones totales. La ausencia de variación podría deberse a que siendo la primera ocasión en que los responsables académicos asumen la responsabilidad de valorar a los profesores en relación a su labor docente o les faltan criterios de valoración o no han querido arriesgarse con valoraciones negativas. A pesar de ello, se considera indispensable incluir la valoración del responsable académico en proyectos de evaluación de la actividad docente que pretendan ser integradores.

En síntesis, creemos que un modelo como el presente, basado en la recogida y análisis de ambos tipos de información, puede facilitar la identificación de “buenas prácticas docentes” que le pueden ser señaladas al participante como tal, para que se mantenga en su uso e incluso puedan ser generalizables y ofrecidas como modelo a los pares, lo que supondría un valor añadido de este tipo de programas.

Una vez aplicado y analizado el modelo, sería conveniente buscar la simplificación, seleccionando los criterios e indicadores que tienen más peso en su definición y reagrupando aquellos que, según los componentes extraídos, saturan de forma conjunta para explicar la varianza. Por otra parte un desafío importante para que estos modelos de evaluación se consoliden tiene que ver con la definición de criterios empíricos externos que lo soporten, que avalen la relación entre la puntuación obtenida cuando los diversos agentes opinan y los resultados que obtienen los estudiantes, su éxito académico, las tasas de éxito, de abandono o de inserción laboral. El trabajo que hemos realizado tiene un alcance limitado, ya que sólo se refiere a una pequeña muestra, será necesario acumular más datos para ver realmente las aportaciones y alcance de estas propuestas de evaluación integral de la actividad docente.

## **Referencias**

- Barth, M.M. (2008): Deciphering student evaluations of teaching: A factor analysis approach. *Journal of Education for Business*, 84(1), 40-46.
- Drews, D.R.-Burroughs, W.J.-Nokovich, D. (1987): Teacher self-ratings as a validity criterion for student evaluations. *Teaching of Psychology*, 14(1), 23-25
- Felton, J.-Mitchell, J.-Stinson, M. (2004): Web-based student evaluations of professors: The relations between perceived quality, easiness and sexiness. *Assessment and Evaluation in Higher Education*, 29(1), 91-108.
- Gamliel, E.-Davidovitz, L. (2005): Online versus traditional teaching evaluation: Mode can matter. *Assessment y Evaluation in Higher Education*, 30(6), 581-592.
- Gigliotti, R.J.-Buchtel, F.S. (1990): Attributional bias and course evaluations. *Journal of Educational Psychology*, 82(2), 341-351.
- Greenwald, A.G. (1997): Validity concerns and usefulness of student ratings of instruction. *The American Psychologist*, 52(11), 1182-1186.
- Greenwald, A.G.-Gillmore, G.M. (1997): Grading leniency is a removable contaminant of student ratings. *The American Psychologist*, 52(11), 1209-1217.
- Griffin, B.W. (2001): Instructor reputation and student ratings of instruction. *Contemporary Educational Psychology*, 26(4), 534-552.
- Griffin, B.W. (2004): Grading leniency, grade discrepancy, and student ratings of instruction. *Contemporary Educational Psychology*, 29(4), 410-425.
- Hobson, S.M.-Talbot, D.M. (2001): Understanding student evaluations. *College Teaching*, 49(1), 26-31.
- Isely, P.-Singh, H. (2005): Do higher grades lead to favorable student evaluations? *Journal of Economic Education*, 36(1), 29-42.
- Kwan, K. (1999): How fair are student ratings in assessing the teaching performance of university teachers? *Assessment y Evaluation in Higher Education*, 24(2), 181-196.

- MacAlpine, M. (2001): An attempt to evaluate teaching quality: One department's story. *Assessment and Evaluation in Higher Education*, 26(6), 563-578.
- Marsh, H.W. (1980): The influence of student, course, and instructor characteristics in evaluations of university teaching. *American Educational Research Journal*, 17(2), 219-237.
- Marsh, H.W.-Roche, L.A. (1997): Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *The American Psychologist*, 52(11), 1187-1197.
- Marsh, H.W.-Roche, L.A. (2000): Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders? *Journal of Educational Psychology*, 92(1), 202-228.
- Miller, A.H. (1988): Student assessment of teaching in higher education. *Higher Education*, 17(1), 3-15.
- Min, Q.-Baozhi, S. (1998): Factors affecting students' classroom teaching evaluations. *Teaching & Learning in Medicine*, 10(1), 12-15.
- Moore, T. (2006): Teacher evaluations and grades: Additional evidence. *Journal of American Academy of Business, Cambridge*, 9(2), 58-62.
- Pfeffer, J.-Fong, C.T. (2002): The end of business schools? less success than meets the eye. *Academy of Management Learning & Education*, 1(1), 78-95.
- Rao, N. (1995): The oh no! syndrome: A language expectation model of undergraduate negative reactions toward foreign teaching assistants. Trabajo presentado en el *79th Annual Meeting of the International Communication Association*, Mayo, Nuevo México.
- Stapleton, R.J.-Morkison, G. (2001): Optimizing the fairness of student evaluations: A study of correlations between instructor excellence, study production, learning production, and expected grades. *Journal of Management Education*, 25(3), 269-291.
- Wachtel, H.K. (1998): Student evaluation of college teaching effectiveness: A brief review. *Assessment and Evaluation in Higher Education*, 23(2), 191-211.
- West, R. F. (1988): The short-term stability of student ratings of instruction in medical school. *Medical Education*, 22, 104-112.