

FACTORES MODULADORES DE LA PERCEPCIÓN DE LA CALIDAD DOCENTE

[Modulating factors in the perception of teaching quality]

por

[Article record](#)

[About authors](#)

[HTML format](#)

Casero, Antonio (a.casero@uib.es)

[Ficha del artículo](#)

[Sobre los autores](#)

[Formato HTML](#)

Abstract

This article examines the existing research findings in relation to the research of the factors that are presumably involved in university students' evaluations of the teaching quality they receive, possibly threatening the validity of the construct. The review is organized around the three implicated sources in the perceptive process of the student, being the student himself, the professor, and the context. The review concludes that, despite some minor effects, the factors under analysis do not represent a substantial threat to the system's validity. The unproductive debate that has gone on for decades about guarantees in the evaluation of teaching staff has largely been used to cover up the absence of a desire to institutionalize this kind of evaluation system.

Keywords

Student evaluation of Teacher Performance, Teaching Quality, Effective Teaching.

Resumen

El presente artículo tiene como objetivo presentar una revisión sobre los hallazgos existentes en relación a la investigación de los factores supuestamente implicados en la valoración que el alumnado universitario realiza sobre la tarea docente de sus profesores, pudiendo suponer una amenaza a la validez de dicho constructo. La revisión se presenta organizada en torno a las tres fuentes implicadas en el proceso perceptivo del alumno, siendo éstas el propio alumno, el profesor y el contexto. La revisión concluye que, a pesar de efectos menores, los factores analizados no representan una amenaza substancial a la validez del sistema. El debate infructuoso que se ha producido durante décadas sobre las garantías en la valoración del personal docente ha sido usado en gran parte para encubrir la falta de deseo de institucionalizar este tipo de sistemas de evaluación.

Descriptores

Evaluación del desempeño docente, Calidad docente, Efectividad docente.

La encuesta de opinión que se aplica al alumnado universitario, para evaluar la eficacia de sus docentes, es hoy una octogenaria cuya salud sigue siendo cuestionada por amplios sectores del profesorado, aquellos en los que existen concepciones muy arraigadas sobre determinados factores que sesgarían o influirían indebidamente las opiniones que los alumnos proporcionan sobre la docencia

de sus profesores. Este ámbito, el de los atentados contra la validez de los cuestionarios de evaluación de la docencia, ha sido prolijo en cuanto a producción científica, concluyendo que ninguno de los supuestos factores de sesgo ha demostrado un efecto sustancial, si bien se ha detectado algún que otro débil impacto (Apodaca y Rodríguez, 1999).

En este trabajo nos proponemos presentar una revisión de las investigaciones que han tenido como objeto de estudio las variables bajo sospecha, y que no nos atrevemos a denominar sesgos, ya que para merecer al calificativo de «fuentes de error» deben tener una influencia externa sustancialmente relacionada con las valoraciones y relativamente no relacionada con otros indicadores de la calidad docente. O como indica Villa «para poder probar que una característica invalida la evaluación no basta una correlación significativa, ni siquiera si es plausible una relación causal; hace falta además que dicha característica no tenga relación positiva con otros indicadores de eficacia docente» (Villa, 1985, 84).

Los elementos básicos de la relación aquí estudiada son: el alumno y el profesor. El segundo lleva a cabo una actuación que es valorada de una u otra forma por el primero. En esta relación no podemos obviar el escenario, tercer elemento que modula la relación. De esta manera, entendemos que los juicios emitidos por el alumno son producto de su proceso perceptivo, estimulado por la conducta del objeto percibido –el profesor–. Este proceso transcurre en un escenario –con sus propias características espaciales y temporales–. Así, estos tres elementos configuran las fuentes de variación del proceso perceptivo, tal como postula la Psicología Social en sus estudios sobre la percepción de personas.

El esquema que utilizaremos en esta exposición consistirá en la adscripción de las variables a cada una de las tres fuentes implicadas en el proceso perceptivo del alumno, siendo éstas el propio alumno, el profesor y el contexto.

VARIABLES ASOCIADAS AL ALUMNO

Vamos a recoger aquellas variables pertenecientes al alumno que han sido objeto de examen en la investigación de los cuestionarios de evaluación de la docencia como posibles moduladoras de la respuesta.

SEXO

Aparicio, Tejedor y Sanmartín (1982), citan una serie de estudios norteamericanos que arrojan resultados contradictorios. Unos autores no encuentran diferencias entre alumnos y alumnas, otros indican que las alumnas son más críticas que los alumnos y éstas suelen valorar mejor a sus profesoras y otros que las alumnas valoran mejor que los alumnos a sus profesores en algunos ítems.

Aleamoni y Hexner (1980) citan referencias similares, y Marsh (1987) llega a conclusiones parecidas.

Fernández y Mateo (1997) llevaron a cabo un análisis de la varianza con tres factores: género del alumnado, género del profesorado y Centro –Facultad de Ciencias, Facultad de Ciencias Sociales y Escuela Técnica–. En el diseño 2 x 2 x 3 fueron utilizadas como variables dependientes los dos factores extraídos, mediante análisis factorial exploratorio y confirmatorio, de la versión reducida del Complutense *University Teachers Evaluation Questionnaire* –CUTEQ-R–: competencia docente y motivación y habilidad de interacción. Los resultados del ANOVA muestran tres componentes estadísticamente significativos: el género del alumno, presentando efecto tanto sobre la dimensión de competencia docente como sobre la dimensión de motivación y habilidad de interacción –las alumnas valoran mejor al profesorado–, y el Centro sobre la dimensión de motivación y habilidad de interacción –mejores evaluaciones desde la Facultad de Ciencias Sociales, seguidas de la Escuela técnica, y a más distancia la Facultad de Ciencias–. El tamaño de los efectos fue verdaderamente bajo –0.0078, 0.0098 y 0.0072, respectivamente, fueron los índices omega cuadrado hallados–. En resumen, tal y como indican los autores, el efecto del género del alumnado sobre las evaluaciones que éste realiza sobre la calidad de sus docentes es muy débil o casi inexistente.

Expectativas

Aunque es más conocido el efecto de las expectativas del profesor sobre el alumno, se sabe que las expectativas de los alumnos también pueden influir sobre su propio comportamiento y sobre el del profesor. Feldman y Prohaska (1979), realizaron dos experimentos:

En el primero de ellos, un profesor dio una clase a un grupo de universitarios. A la mitad de ellos se les dijo que el profesor era muy competente y a la otra mitad, lo contrario. Se encontraron diferencias significativas en las actitudes, rendimiento y conducta no verbal entre los grupos. Los alumnos que esperaban un profesor competente: calificaron la lección como fácil e interesante, evaluaron al profesor como más competente, inteligente y entusiasta; obtuvieron puntuaciones superiores en un test de rendimiento y mostraron una conducta verbal diferente, caracterizada por inclinarse más hacia el profesor y mirarle con más frecuencia.

En el segundo experimento, se pidió a un grupo de universitarios que mostraran la conducta no verbal anteriormente descrita –inclinarse la cabeza y mirar hacia el profesor– y otro grupo que mostraran la conducta contraria cuando recibieron una clase de un profesor –en realidad otro alumno que hacía de profesor–. Los sujetos que habían dado la clase a los alumnos de conducta no verbal positiva: se sintieron mejor durante la misma, se percibieron a sí mismos más competentes, percibieron a los alumnos como más entusiastas y agradables, fueron más eficaces y mostraron una conducta no verbal diferente caracterizada por inclinarse menos hacia los alumnos que los otros sujetos.

Teacher Folklore

Estrechamente ligado a la idea de expectativas, existen una serie de estudios bajo la fórmula de «*Teacher Folklore*». Estos trabajos analizan el efecto de la opinión que los alumnos tienen al comenzar el curso y antes de conocer al profesor sobre las valoraciones

que éste recibe al final del mismo. La idea previa del profesor fundamentalmente viene dada por los alumnos de cursos superiores. Los aspectos de reputación suelen incluirse también en esta variable, denominada en ocasiones «actitudes de entrada».

Aleamoni et al. (1972) estudiaron el *Teacher Folklore* operativizado en contenido del curso, carga de trabajo, estilo de profesor... y a través de correlaciones pretest-postest concluyeron que el «*teacher folklore*» no contribuye significativamente a las valoraciones del curso.

En otro trabajo, realizado por Barke, Tollefson y Tracy (1983), se intentó relacionar las actitudes al comienzo y las valoraciones al final. Para ello seleccionaron, de forma aleatoria y estratificada por edad y rango, a 75 profesores de un total de 500. Estos profesores fueron evaluados por 789 alumnos que contestaron al pretest y al postest. Los resultados indican que sólo una minoría de los estudiantes expresó expectativas sobre el docente y el curso, y en estos pocos casos las valoraciones del profesor son moderadamente predecibles. De manera que «los profesores pueden ser víctimas –o beneficiarios– de sus reputaciones» (Barke, Tollefson y Tracy, 1983, 83). De acuerdo con Marsh (1984), las expectativas pueden afectar a las valoraciones pero para la mayoría de estudiantes las valoraciones reflejan sus juicios sobre la calidad de la enseñanza.

Interés previo por el tema

Los estudios disponibles al respecto han sido realizados en relación al *SEEQ*, por Marsh y otros autores que, a su vez, revisaron estudios previos. El interés previo por la materia correlacionó con la dimensión aprendizaje/valor, $r = 0.40$. «Un alto interés del estudiante por el tema aparentemente crea un entorno de aprendizaje más favorable y facilita la docencia efectiva, y este efecto se refleja en las valoraciones estudiantiles» (Marsh, 1987, 315).

Los resultados no aclaran si un mismo docente, impartiendo materias sin interés previo para el alumnado, obtendría las mismas valoraciones.

Personalidad de los estudiantes

En un trabajo de Abrami y D'Apollonia (1990) se presentan tres estudios que examinan la relación entre características de la personalidad del estudiante, valoraciones del profesor y rendimiento del estudiante. En el primer estudio 388 estudiantes se autovaloran en la *Adjective Check List* –ACL– (Gough, 1979), para después ver uno de cuatro videos que varían en la expresividad del profesor –baja, alta– y el contenido de la explicación –bajo, alto–, posteriormente completar una evaluación del profesor del vídeo y un test sobre la explicación; finalmente completan el ACL para el profesor. El segundo estudio fue similar excepto en que los estudiantes, 87 en este caso, fueron expuestos a dos videos separados por una semana. En el tercer estudio los estudiantes, 108, de cinco clases de Introducción a la Psicología completaron el ACL para sí mismos, evaluaron la docencia de su profesor, completaron un test sobre el material común del curso y completaron el ACL para su profesor. Los hallazgos sugieren cuatro conclusiones relevantes para la evaluación sumativa de la docencia (Abrami y D'Apollonia, 1990, 111):

- a) No parecen existir relaciones significativas ni consistentes entre valoraciones y personalidad de los estudiantes.
- b) Las características de personalidad de los profesores, tal como son percibidos por los estudiantes, están relacionadas con las valoraciones de la efectividad docente.
- c) Las valoraciones predicen rendimiento producido por el profesor justamente mejor para clases en las que difieren las características de personalidad de los alumnos matriculados.

- d) Los efectos del profesor sobre las valoraciones aparecen significativamente mayores que estos efectos sobre el rendimiento.

Estilos evaluativos

Un interesante trabajo de Castro (1996) analiza la existencia de lo que denomina estilos evaluativos, para referirse a patrones estables de evaluación por parte del alumnado, es decir, un estilo evaluador permanente que con independencia de lo evaluado, siempre, o casi siempre, van a ver la realidad evaluada de la misma manera. El autor sustenta sus hipótesis sobre la base de los estilos cognitivos y la teoría de los moldes cognitivos-afectivos (Hernández, 1991).

El estudio fue realizado sobre 6.040 alumnos de la Universidad de las Palmas de Gran Canaria, a los que se les aplicaron cuatro instrumentos:

- Cuestionario de evaluación del profesorado: 35 ítems, arrojando 6 factores que explican el 53% de la varianza. En total se realizaron 23.816 evaluaciones, donde cada alumno evaluó una media de cuatro profesores.
- Cuestionario de evaluación de los elementos contextuales: compuesto por 14 ítems en el que se evalúa el grado de satisfacción de los alumnos con una serie de elementos que conforman su contexto académico más inmediato –relación entre alumnos, ratios, servicios de la Universidad...–. El instrumento arrojó cuatro factores que explicaron el 47.7% de la varianza.
- Cuestionario de atribución causal de la mala calidad de la enseñanza universitaria: compuesto por 14 ítems relativos a posibles causas relacionadas con la explicación de una mala calidad de enseñanza. El instrumento arrojó cuatro factores que explicaron el 53.1% de la varianza.
- Cuestionario de atribución causal del fracaso académico de los alumnos: compuesto por 11 ítems relativos a posibles causas del fracaso académico. El instrumento

arrojó tres factores que explican el 48.3% de la varianza.

Las conclusiones del estudio revelan, entre otros aspectos, que «ha quedado demostrada la eficacia del alumnado como evaluadores objetivos en su mayoría, no sesgados, fiables, válidos y con amplia capacidad para discriminar entre distintos profesores y a su vez entre estos y otras situaciones», y que «existe un estilo evaluador de alumno respecto al profesorado que permite clasificar a una minoría de alumnos como:

- A. *Críticos*, que valoran peor a sus profesores, se muestran más insatisfechos con su contexto académico y achacan más la responsabilidad de la mala calidad de la enseñanza y el fracaso académico a los elementos más asociados al profesorado.
- B. *Benévolos*, que se comportan de manera contraria a los anteriores.
- C. Y una mayoría de alumnos *ecuánimes* que se sitúan en sus valoraciones entre los dos grupos anteriores» (Castro, 1996, 471).

El autor finaliza aconsejando el uso de ítems control para prever el estilo evaluador general del aula, para saber si predomina algún estilo extremo o, por el contrario, están igualados.

Posteriormente, Etopa (2003) analizando los resultados de las encuestas de la Universidad de las Palmas de Gran Canaria comparó éstos bajo dos indicadores de tendencia central: la media aritmética y la media recortada al 5%. Éste último es más robusto que el primero al eliminar del cálculo el 10% extremo. Los resultados le llevaron a concluir que «el hecho de eliminar las evaluaciones en las que se emite un juicio sesgado – crítico-benévolos – no hace aumentar significativamente la media y refuerza que los estudiantes de la ULPGC son evaluadores válidos y fiables» (Etopa, 2003, 269).

Notas esperadas y/o recibidas

Peterson y Cooper (1980) indican que algunos estudios han encontrado una relación positiva entre la calificación esperada o recibida por un estudiante y las valoraciones del profesor que califica. Existe la creencia de que los alumnos premian a los profesores que les han calificado positivamente y castigan a los que les han calificado negativamente. Estos autores realizaron un estudio con dos grupos de profesores, uno que calificaba a los alumnos y otro que les daba o no créditos y, comparando las valoraciones de los dos grupos, llegaron a la conclusión de que eran similares.

Para Marsh (1980), la variable contextual que razonablemente puede suponer un sesgo para las valoraciones es la calificación esperada, aunque esta interpretación esté sujeta a explicaciones alternativas. Por parte del alumno, el interés previo por el tema puede estar en la base de una mejor experiencia de aprendizaje que propicie mejores calificaciones y, por parte del profesor, la mayor efectividad docente puede estar causando mejor aprendizaje en los alumnos y calificaciones más altas. También, al estar correlacionada positivamente, la demanda de trabajo con las valoraciones, puede suceder que los profesores que demandan trabajo a los alumnos sean valorados más positivamente y, a su vez, los alumnos obtienen mayor rendimiento – mejores calificaciones–.

Aleamoni y Hexner (1980) citan gran cantidad de estudios mostrando la controversia en torno a las valoraciones estudiantiles y las calificaciones esperadas o recibidas. La tendencia general es que los alumnos tienden a valorar más positivamente a los cursos y profesores cuando reciben o esperan recibir buenas calificaciones. Ofrecen una mediana de las correlaciones de 0.14 aproximadamente, con una media y desviación de 0.18 y 0.16 respectivamente.

Aparicio, Tejedor y Sanmartín (1982), en la revisión que hacen de los estudios sobre el

tema, indican que, en general, se ha encontrado una correlación positiva entre calificaciones y valoraciones. No obstante, las correlaciones obtenidas oscilan entre -0.75 y 0.75. Esta falta de acuerdo la atribuyen a problemas de orden metodológico.

Según Marsh (1984), la relación positiva entre medias de calificaciones esperadas y valoraciones puede plantear tres hipótesis:

- Benevolencia en las calificaciones: los profesores dan calificaciones más altas y reciben valoraciones más altas que las merecidas.
- Validez: mejores calificaciones esperadas reflejan un mejor aprendizaje de los estudiantes y una correlación positiva fundamenta la validez de las valoraciones estudiantiles.
- Características de los estudiantes: preexisten variables pronóstico de los estudiantes tales como, el interés previo por la materia que afectan al aprendizaje del estudiante, a las calificaciones y a la efectividad docente de manera que el efecto de las calificaciones esperadas es espurio.

Para finalizar con este punto presentamos los resultados de un trabajo llevado a cabo en la Universidade da Coruña por De Salvador (1996). El estudio utiliza como muestra 345 asignaturas, correspondientes a 16 titulaciones –11 centros– y con más de 20 cuestionarios contestados en cada asignatura. De cada asignatura se ha obtenido la valoración promedio de cada una de las tres dimensiones que componen el cuestionario de evaluación de la docencia por los alumnos en esta Universidad, y la valoración realizada por el profesor en las calificaciones dadas a los alumnos –las notas–. A continuación presentamos un resumen de los resultados:

- Se ha observado una manifiesta relación entre las calificaciones dadas por el profesor y las valoraciones de los alumnos en el cuestionario de evaluación docente. Este hecho hace pensar al autor en un efecto halo de éstas: calificaciones altas/bajas en

una asignatura tendría cierto correlato en valoraciones altas/bajas de los alumnos para esa asignatura. Las correlaciones halladas entre las calificaciones y las medidas del instrumento son: ítem buen profesor, 0.391; factor metodología didáctica, 0.350; factor relación interacción, 0.482; y factor evaluación, 0.496, todas ellas con significaciones inferiores a 0.001.

- Dicha relación está más mediatizada por las «buenas notas» que por los suspensos.
- El efecto halo se consolida por la congruencia de la relación ateniéndose a los componentes del cuestionario: el sistema de evaluación y la relación con los alumnos son las dimensiones más influenciadas.
- La mayor influencia se observa en las asignaturas que son impartidas por un único profesor de un curso terminal –sobre todo si se trata de una diplomatura–.
- El curso, en el ámbito general, presenta una tendencia cuadrática del efecto con un punto de inflexión máximo en tercero: aumenta moderadamente durante el primer ciclo con disminución, también moderada, en el segundo ciclo.

El autor concluye que las calificaciones pueden estar distorsionando la evaluación docente por los alumnos, entendiendo que el principio de coherencia teórica «un buen profesor da como resultado buenas calificaciones de sus alumnos», no puede ser interpretado dados los resultados obtenidos, «ahora bien, consideramos que este estudio está circunscrito a un único año de aplicación del cuestionario y que en todo caso debe ser ampliado a años sucesivos, no sólo con tratamiento transversal sino también longitudinal» (De Salvador, 1996, 125).

Variables asociadas al profesor

A continuación presentamos las variables asociadas al profesor que han sido objeto de estudio en la investigación de los cuestionarios de evaluación de la docencia por los

alumnos como posibles fuentes que modulan las respuestas.

Edad, sexo y aspecto físico

En un estudio de Goebel y Cashen (1979) se comprobó que los alumnos tienden a evaluar peor a los profesores de más edad y los menos atractivos. El efecto sexo no llegó a ser estadísticamente significativo. El análisis de las interacciones mostró que las mujeres no atractivas de mediana edad y los hombres no atractivos de mayor edad tendían a ser peor evaluados.

Villa expresa en este sentido que «los profesores más viejos y menos atractivos empiezan con una desventaja. Ello refleja los estereotipos de la sociedad, ya que los profesores son evaluados en base a características distintas de las profesionales; la validez de las evaluaciones estudiantiles, al hacer juicios de la competencia profesional, queda, según los autores, seriamente cuestionada» (Villa, 1985b, 45).

Pozo, Reboloso y Fernández (2000) han realizado un estudio que pretende determinar cuales son las características que definen a un «profesor ideal». Para ello aplican, a 2.221 estudiantes de la Universidad de Almería, una escala de diferencial semántico compuesta por 29 adjetivos bipolares. Los resultados muestran cómo el físico de los profesores «atractivos o interesantes» es bastante significativo. Este hallazgo lleva a los autores a hipotetizar que los aspectos emocionales tienen una gran influencia en la actitud de los estudiantes hacia sus profesores y, en consecuencia, en la conceptualización que de ellos hacen.

Recientemente, un estudio cualitativo con similar objetivo que el anterior fue llevado a cabo por el autor de este trabajo. En él se utilizaron entrevistas cualitativas y grupos de discusión. El análisis de contenido dio lugar a una subcategoría emergente que se denominó aspecto físico, en relación a las declaraciones de unos pocos alumnos sobre el

efecto positivo que sobre la idea global del docente puede tener su imagen. Esta subcategoría fue la que menos producción verbal tuvo asociada, razón por la que su presencia en la estructura general de nodos fue cuestionada por su muy baja representatividad (Casero, 2010).

Personalidad

La revisión y síntesis llevada a cabo por Feldman (1986) sobre los trabajos que relacionan valoraciones del estudiante y rasgos o categorías de personalidad obtenidas de autoinformes y heteroevaluaciones –estudiantes, colegas– muestran lo siguiente:

- Autoinformes: correlaciones positivas sólo en autoimagen positiva/ autoestima $-r = 0.30-$ y energía/entusiasmo $-r = 0.27-$. La correlación media entre valoraciones estudiantiles de la efectividad docente y cada una de las 12 categorías restantes eran en 0.15 o menor.
- Cuando la personalidad fue valorada por observadores externos –colegas y alumnos– las correlaciones fueron mucho más altas, y el promedio de éstas entre valoraciones estudiantiles y los 14 rasgos de personalidad fue de 0.30 a 0.60.

El autor ofrece varias explicaciones posibles para estas diferencias:

- La personalidad del profesor percibida por los estudiantes puede verse afectada por los mismos sesgos que afectan a las valoraciones de la efectividad docente.
- La personalidad inferida por los colegas puede basarse en parte sobre informaciones de los estudiantes.
- La personalidad inferida por los estudiantes y colegas puede estar basada, en parte, en las percepciones de la efectividad docente más que las de la personalidad –o viceversa–.
- La personalidad inferida de autoinformes puede ser más o menos válida, o estar más o menos sesgada, que la personalidad inferida por estudiantes y colegas. La personalidad

inferida por estudiantes, y quizás por colegas, puede limitarse a un aspecto situacional específico, mientras que las medidas de la personalidad en base a autoinformes son medidas más generales y no centradas en un contexto específico.

Para Marsh (1987), la revisión de Feldman sugiere que hay una relación entre evaluaciones de los estudiantes de un profesor y al menos algunos aspectos de la personalidad del profesor, pero no nos indica que sea una fuente válida de influencia o de sesgo.

Para Díaz-Aguado (1987), la mejor forma de considerar la influencia de las características individuales del profesor –como tipo de personalidad o estilo cognitivo– sobre las evaluaciones de la docencia es en la interacción con los mismos rasgos del alumno.

Impartición del curso en varias ocasiones

En este apartado se trata de determinar si el hecho de que un profesor haya impartido más de una vez un curso hace que aumenten las valoraciones. Marsh llegó a las conclusiones siguientes: las correlaciones entre valoraciones estudiantiles del mismo curso impartido por el mismo profesor en diferentes ocasiones fueron altas – $r = 0.71$ –. Para cada par de cursos, 341 pares, el más favorablemente evaluado tiende a ser:

- a) Uno en el que los estudiantes esperan calificaciones más altas –y presumiblemente mayor aprendizaje–.
- b) Uno en el que los estudiantes perciben que se les exige más trabajo.
- c) Uno que fue impartido después de que el profesor hubiera impartido el curso al menos una vez –y presumiblemente mejorado como consecuencia de su experiencia–.

En resumen, el autor manifiesta «este estudio relaciona diferencias en las valoraciones del mismo profesor que imparte el mismo curso en dos ocasiones diferentes con diferencias en las características contextuales de los dos cursos. En cada par de cursos, la valoración más alta ofrecida estuvo correlacionada con

más altos niveles de carga de trabajo/dificultad; mayores calificaciones esperadas –y presumiblemente mayor maestría del curso–; y que el profesor haya impartido el curso al menos una vez con anterioridad. Ninguno de los hallazgos es consistente con la noción de que las evaluaciones estudiantiles están sesgadas por las características contextuales, y las conclusiones soportan explicaciones alternativas a estas relaciones» (Marsh, 1982 b, 496).

Voluntarismo a ser evaluado

En un trabajo, llevado a cabo por Howard y Bray (1979), se descubrió que un grupo de voluntarios –profesores que continuaron usando las valoraciones estudiantiles tras un primer curso obligatorio– obtuvieron valoraciones más altas que los no voluntarios.

Basados en este estudio, Cashin y Perrin (1983) analizan las valoraciones estudiantiles mediante el *Instructional Development and Effectiveness Assesment* –IDEA– en 13.063 clases de algunos campos académicos e instituciones. Las clases estaban divididas en tres grupos: voluntarios, en los que la decisión de evaluar fue enteramente del profesor; intermedios, la evaluación fue exigida, pero el profesor eligió la clase; y no voluntarios, en los que el profesor se le exigió la evaluación por clase. Las diferencias encontradas entre los tres grupos fueron estadísticamente significativas para 26 de los 39 ítems del IDEA –probablemente a causa del amplio número de casos–. Basados en análisis de omega cuadrado ninguna de estas diferencias tienen significación práctica –los tamaños del efecto no alcanzaron siquiera el 1% de la varianza–. Se concluyó que la voluntariedad de la evaluación no necesita ser tomada en consideración cuando se usan instrumentos de recogida de datos grandes, multinstitucionales y comparativos.

Quizá resulte conveniente plantearse la hipótesis de que a mayor efectividad valoraciones más altas, más preocupación por mejorar y lógicamente mayor interés –

voluntariedad– en conocer las opiniones de los alumnos sobre la efectividad docente.

Rango y experiencia

Otro tema estudiado es el del rango del profesor como posible influencia en las valoraciones estudiantiles. El tema ha sido estudiado por varios autores. Entre los trabajos destaca la revisión realizada por Feldman (1983) donde indica que, en general, los profesores asistentes reciben calificaciones más bajas que otros profesores en la mayoría de las dimensiones del *SEEQ* y en algunos ítems, aunque pueden recibir valoraciones más altas en las dimensiones «relación con el alumno individual» e «interacción con el grupo». En la mayoría de los estudios que revisa este autor encuentra un efecto no significativo del rango del profesor en valoraciones globales, las relaciones significativas tienden a ser positivas. Tampoco está relacionado significativamente con las dimensiones de las valoraciones en la mayoría de los estudios, cuando son positivas lo más probable es que lo sean con «conocimientos» y «expansividad intelectual», mientras que la relaciones negativas aparecen con «animación de la discusión» y «apertura y preocupación por los estudiantes».

Respecto a la experiencia, y muy relacionado con la edad, Braskamp et al. (1985) encontraron que las valoraciones estudiantiles pueden incrementarse durante los 10 primeros años de docencia, pero decrecen un poco tras ellos.

El paradigma del Dr. Fox

El efecto del Dr. Fox, también conocido como *seducción académica*, se define como la influencia predominante de la expresividad del profesor. Este fenómeno suele asociarse con que el entusiasmo del docente puede seducir a los alumnos para que otorguen evaluaciones favorables a su desempeño, aún cuando la instrucción carezca de contenidos relevantes.

El estudio original sobre el efecto del Dr. Fox fue realizado por Naftulin, Ware y Donnelly, en 1973. Los autores contrataron a un actor profesional para dar una serie de conferencias a tres grupos de profesores y estudiantes de medicina; el actor fue presentado como el «Doctor Myron L. Fox». Éste fue previamente entrenado para que empleara neologismos, ambigüedades, contradicciones y frases sin sentido en sus conferencias. En resumen, el Dr. Fox dio una serie de conferencias entretenidas y con poco o ningún contenido. Tanto alumnos como profesores evaluaron como favorable el desempeño del instructor.

Ware y Willians (1975, 1977) y Willians y Ware (1976, 1977) desarrollaron el estándar del paradigma del Dr. Fox con series de seis explicaciones, todas ellas presentadas por un actor profesional grabadas en vídeo. Cada explicación representaba uno de los tres niveles de contenido del curso y uno de los dos niveles de expresividad expositiva. Los estudiantes rellenaban un cuestionario de valoración de varios ítems y un test de rendimiento. Los resultados concluyeron que las diferencias en expresividad explican consistentemente mucha más varianza en las valoraciones estudiantiles que diferencias en el contenido.

Marsh y Ware (1982) en un reanálisis de los estudios anteriores llega a la conclusión de que la manipulación de la expresividad del profesor sólo afecta a las valoraciones del «entusiasmo» del profesor, el factor más lógicamente relacionado con ella, y la amplitud de contenido afectó significativamente a las valoraciones del factor «conocimientos del profesor» y «organización/claridad», los factores más lógicamente relacionados con la expresividad. Marsh indica que «un efecto que ha sido interpretado como un sesgo de las valoraciones estudiantiles se interpreta más apropiadamente como soporte para su validez con respecto a algún componente de la enseñanza efectiva» (Marsh, 1987, 333).

Abrami, Leventhal y Perry (1982) realizaron un meta-análisis de todos los estudios conocidos sobre el tema y concluyeron que las manipulaciones de la expresividad tienen un efecto sustancial en las valoraciones estudiantiles sobre todo y un efecto pequeño en el rendimiento, mientras que las manipulaciones del contenido tienen un efecto sustancial en el rendimiento y un efecto pequeño en las valoraciones. Concluyen que a pesar de que la expresividad interactuó con la manipulación de contenido y un grupo de otras variables examinadas en los estudios Dr. Fox, ninguna de las interacciones dio cuenta de más del 5% de la varianza de las valoraciones estudiantiles.

Variables asociadas al contexto

En este grupo de variables recogemos aquellas que no pueden asociarse directamente ni al alumno ni al profesor, como son las relativas a las características de la asignatura –hora en que ésta se recibe, tipo y volumen de trabajo–, el tamaño de la clase y el momento en el que se responde al cuestionario de evaluación de la docencia.

Hora en que se imparte la clase

Algunos profesores opinan que las clases impartidas a última hora de la mañana o a última hora de la tarde pueden ser peor evaluadas porque muchos alumnos se van. Una cosa es que los alumnos se vayan y otra que las valoraciones no se ajusten a realidad al ser sesgadas por este factor. Como mucho se dispondría, en estos casos, de un número menor de alumnos con un juicio formado que les permita evaluar al docente. La verdad es que no hay muchos estudios al respecto, tal como indican Aparicio, Tejedor y Sanmartín (1982).

Cranton y Smith (1986) estudiaron las diferencias entre cursos diurnos y nocturnos, hallando una diferencia muy pequeña bajo un análisis multivariado. Sin embargo, en el estudio por separado –análisis univariante– no encontraron diferencias significativas en

las valoraciones de la docencia y sí en la valoración de la cantidad aprendizaje, la importancia del mismo y la valoración global.

Tipo de asignatura

En este punto recogemos las consideraciones y estudios sobre la importancia de la asignatura en el plan de estudios, su carácter de obligatoria u optativa, y su carácter como disciplina –ciencias, letras...–.

Según indican Aparicio, Tejedor y Sanmartín (1982), no se encontró relación entre la importancia de la asignatura en los estudios y las valoraciones estudiantiles. Estos autores señalan que los resultados de los estudios que relacionan las valoraciones y el carácter de optativa/obligatoria de las asignaturas son contradictorios. Algunos estudios afirman que los alumnos tienden a valorar mejor a los profesores de asignaturas optativas y en otros no encontraron diferencias.

En otro orden, diversos autores han relacionado las valoraciones con el tipo de disciplina académica. Feldman (1978) en una revisión de los estudios que comparaban las valoraciones entre disciplinas encontró que las valoraciones eran algo más altas que el promedio en inglés, humanidades, artes, lenguaje y educación. Y algo más bajas en ciencias sociales, ciencias físicas, matemáticas e ingenierías. Estando en torno a la media en ciencias biológicas.

Centra y Creech (1976) en un amplio estudio –100 instituciones– encontraron que las valoraciones eran más altas en humanidades y más bajas en ciencias.

Neumann y Neumann (1985) en base a la clasificación de áreas académicas de Biglan (1973): a) dura/blanda; b) aplicada/pura; y c) viva/muerta, indican que las valoraciones podrían ser más altas en disciplinas blandas, puras y muertas y, por tanto, las comparaciones deben hacerse entre profesores de áreas parecidas. Sin embargo, este estudio se realizó en una sola institución por lo que su generalizabilidad debería ser probada.

Los resultados de estos estudios pueden tener mayor trascendencia e importancia para la evaluación sumativa que para la formativa. Las disciplinas que indica Feldman como mejor valoradas parecen prestarse a una mayor interacción profesor-alumno, al uso de métodos más activos, y en definitiva a la participación del alumnado.

En un estudio llevado a cabo por García Valcárcel (1989), en la Universidad de Cantabria, se aplicaron cuestionarios los alumnos y, tras un análisis de *cluster* de las respuestas promedio por ítem/asignatura, aparecieron dos modelos nítidamente diferenciados:

- Modelo informativo: incluye 84 asignaturas y está caracterizado por los siguientes elementos; memorización de conocimientos, exposición a modo de conferencia, uso de un único libro de texto y empleo de exámenes tradicionales. El profesor informa, el alumno asimila, el alumno demuestra lo que ha asimilado en un examen, y el profesor califica.

- Modelo comunicativo o formativo: 62 asignaturas. Con puntuaciones elevadas en los ítems referentes a objetivos, motivación, interacción y, ocasionalmente, uso de técnicas de enseñanza distintas a la clase magistral.

Carga de trabajo/dificultad de la asignatura

Éste ha sido otro aspecto tenido en cuenta. Entre las razones por las que conviene prestar atención a la presencia de algún efecto en el sentido del título de este punto, se encuentran las indicadas por Ryan, Anderson, y Birchler (1980) cuando exponen que la introducción de la evaluación por los estudiantes en una institución produce un descenso en la exigencia y dificultad de los cursos, en la opinión de que esto podría conducir a valoraciones estudiantiles de los profesores más altas.

Las investigaciones en cuyos cuestionarios se incluye este aspecto, coinciden en que la exigencia y dificultad de los cursos se asocia positivamente con valoraciones más favorables en igualdad de circunstancias en otros aspectos. Para investigar este tema se suele

estudiar el mismo curso impartido por el mismo profesor en ocasiones diferentes. Sin embargo, aunque en las autoevaluaciones del profesor la correlación es menor, sigue manteniendo la misma dirección. Por esta razón no puede decirse que la exigencia/dificultad de la asignatura introduzca un sesgo en las valoraciones estudiantiles (Marsh, 1987).

Tamaño de la clase

En general, los profesores opinan que los colegas que imparten sus clases en grupos pequeños obtienen mejores valoraciones que los que las imparten en cursos más numerosos. Los trabajos realizados al respecto no han encontrado, en la mayoría de los casos, relación alguna. Mientras que en unos pocos ha aparecido una correlación cuadrática. No obstante, los resultados son algo contradictorios.

Aparicio, Tejedor y Sanmartín (1982) citan estudios en los que las valoraciones en grupos pequeños son más elevadas que en grupos numerosos.

Aleamoni y Grahan (1974), entre otros, no han hallado relación significativa entre el tamaño de la clase y la evaluación de la calidad docente por el alumno.

Marsh y sus colegas encontraron que el tamaño de la clase está moderadamente correlacionado con «interacción con el grupo» y «relación con el alumno individualmente» – correlaciones negativas de hasta -0.30–, pero no con otras dimensiones o con las valoraciones globales del profesor o del curso. El autor indica que «la especificidad del efecto del tamaño de la clase en las dimensiones más lógicamente relacionadas con esta variable, y la similitud de resultados basados en valoraciones estudiantiles y autoevaluaciones del profesor apoyan la hipótesis de que este aspecto no es un sesgo para las valoraciones estudiantiles; más aún, el tamaño de la clase ejerce un efecto moderado sobre algunos aspectos de la docencia efectiva, principalmente sobre interacción con el grupo y relación individual,

y estos efectos son exactamente reflejados en las valoraciones» (Marsh, 1987, 314).

Feldman (1978) realizó una amplia revisión hallando resultados consistentes con los ofrecidos por Marsh sobre el *SEEQ*.

Mateo y Fernández (1996), llevaron a cabo un estudio, en la Universidad Complutense de Madrid, sobre un total de 5.959 clases de diferentes tamaños –desde 3 a 498 estudiantes por clase–. Clasificaron los tamaños de acuerdo con las siguientes cinco categorías:

- Muy pequeña: entre 3 y 9 estudiantes
- Pequeña: entre 10 y 29 estudiantes
- Mediana: entre 30 y 59 estudiantes
- Grande: entre 60 y 149 estudiantes
- Muy grande: más de 149 estudiantes

El instrumento utilizado fue el propio de la Universidad –CUTEQ-R– en su versión reducida, donde analizaron las dos dimensiones de «competencia» y «motivación y habilidad de interacción».

El tratamiento estadístico consistió en el análisis de la varianza unifactorial de cada una de las dimensiones bajo el factor anteriormente descrito. Los resultados muestran significación estadística en ambas dimensiones, alcanzando índices t al cuadrado de 0.0299 en la dimensión «competencia», y de 0.0580 en la dimensión «motivación y habilidad de interacción». La tabla siguiente recoge las medias obtenidas.

Tabla 1. Medias obtenidas en las dos dimensiones analizadas según el tamaño del aula

Dimensión Tamaño	Competencia	Motivación y habilidades de interacción
Muy pequeña	5.14	5.82
Pequeña	4.80	4.91
Mediana	4.67	4.74
Grande	4.87	4.75
Muy grande	4.58	4.50

Reproducida a partir de Mateo y Fernández (1996, p. 776)

Los autores concluyen que los resultados dan soporte a la presencia de algún efecto del tamaño del aula sobre las valoraciones del alumnado, citando las coincidencias con los resultados de Feldman y Marsh, y constando una influencia diferencial sobre las dimensiones de la evaluación de la docencia. Este efecto no ha sobrepasado el 5% de la varianza explicada en el mayor de los casos.

Salvador (1990), propone una distinción en la naturaleza de las relaciones entre el tamaño del aula y la valoración del alumno. Señala que, podríamos hablar de «sesgo» cuando simplemente por dar clase a un grupo pequeño se obtengan mejores evaluaciones. No es así cuando el tamaño es un elemento integrante de la efectividad docente, es decir, que la mayor interacción y trato con el alumno que puede producirse genere una mayor efectividad docente y, consecuentemente, mejores valoraciones.

Momento en el que se aplica el cuestionario

En este último punto abordamos la influencia que puede producir sobre las valoraciones el momento del curso en el que éstas son aplicadas.

Feldman (1979) informa de que las valoraciones tienden a ser similares ya sean recogidas en la mitad del periodo escolar o después de completar el curso.

Braskamp et al. (1985) sugieren que las valoraciones recogidas durante el examen final son más bajas, y las de la mitad del curso menos fiables si los alumnos pueden ser identificados.

Marsh y Overall (1980) recogieron valoraciones en la mitad del curso y durante la última semana del curso en su estudio de validez multisección. Ambas estaban altamente correlacionadas, pero los coeficientes de validez de las valoraciones de mitad de curso eran sustancialmente más bajos.

Braskamp et al. (1985) recomiendan que las valoraciones debieran recogerse durante

las dos últimas semanas del curso. Ésta suele ser, aproximadamente, la práctica habitual de las universidades españolas actualmente.

Para no concluir

La revisión realizada sobre la investigación de los «sesgos» que pueden estar afectando a las valoraciones que los estudiantes realizan sobre la calidad de la docencia de sus profesores ha puesto de manifiesto la existencia de algunas variables que presentan, en algunos casos, correlaciones que, si bien alcanzan la significación estadística –resultado lógico trabajando con muestras de gran tamaño–, no suponen un efecto sustancial que atente contra la validez de los instrumentos analizados, tal y como ha sido señalado por varios autores (Marsh, 2007).

Probablemente, los resultados contradictorios detectados entre algunos trabajos sean atribuibles a la falta de coincidencia en los métodos, escalas, unidades de medida utilizados y, en algunos casos, cuestionables posibilidades de representatividad y generalización. A estas faltas de rigor metodológico hay que sumarle el desacierto en la interpretación de algunos resultados. Así, parece lógico pensar que un profesor de mayor rango desarrolle con mayor efectividad su labor docente y, consiguientemente, reciba mejores valoraciones del alumnado ¿es por esto el rango del profesor un sesgo?, o que con tamaños grandes de clase se obtengan peores valoraciones ¿el tamaño de la clase como fuente de error o las condiciones de la enseñanza como factor explicativo de la calidad de la misma? Estos ejemplos, entre otros, ilustran lo que Feldman (1997) denominó los mitos y las medias verdades en la evaluación de la docencia.

Después de cincuenta años de investigación la literatura científica al respecto muestra que, aún con algún débil impacto, la medida de las valoraciones de la docencia por parte del alumnado es fiable y válida. Pocos objetos de estudio, en el ámbito de la evaluación, han sido tan cribados, y como conse-

cuencia de este obsesivo escrutinio goza de más garantías que mucha otra instrumentación de uso extendido. Sin embargo, la investigación de las últimas décadas no ha contribuido especialmente a resolver los flecos detectados en los años ochenta, década de oro en la investigación de las valoraciones de la docencia por parte del alumnado. La producción científica más reciente está orientada hacia la exquisitez psicométrica, en el caso sajón, mientras que la literatura española del último par de décadas parece más preocupada en la finalidad de la evaluación –sumativa y/o formativa–, sin consenso al respecto; mientras autores como De Miguel (1991) aboga por la complementariedad de ambos fines, otros como Escudero (1993) se oponen a esta visión. Por su parte Meliá (1993) y Mateo (2000) proponen procesos mixtos basados en la contextualización y el consenso, y Apodaca y Grad (2002) subrayan que es el fin de la evaluación el que condiciona la visión unidimensional o multidimensional del constructo. En este sentido, tal y como se ha apuntado en otros trabajos (Casero, 2008, 2010), el apoyo mayoritario lo recibe el modelo multidimensional.

Recapitulando, resulta curioso como un objeto de estudio tan escrutado y bien definido siga generando polémica a lo largo de décadas. La cuestión es ¿por qué? Quizá parte de la respuesta, sino toda, venga dada por el punto en el que la evaluación incide, la competencia del docente. Este parece ser un asunto tremendamente delicado, más aún si el juicio sobre dicha competencia queda en manos del alumnado. El sistema de evaluación de la docencia se conforma a partir de una serie de agentes, entre ellos el alumnado, sin embargo en muchas universidades españolas éste ha sido la única fuente de información utilizada. Este hecho junto con los prejuicios personales, el miedo, la desconfianza, la ignorancia y la conveniencia administrativa, entre otras causas, han sostenido el debate improductivo sobre la fiabilidad y validez de las evaluaciones de la docencia, ignorando así los hallazgos de cincuenta

años de investigación (Theall y Feldman, 2007). El problema parece originarse en la falta de valentía, de las instituciones principalmente. Las universidades deben afrontar la creación de sistemas de evaluación totales, desarrollar programas de formación del profesorado coherentes con unos estándares de calidad docente y fomentar una cultura de la calidad. Tal vez así, con la implicación institucional, puedan llevarse a cabo mejores investigaciones que, además de la clara mejora del sistema universitario español, redunden también, en el destierro absoluto de los mitos y las medias verdades.

Referencias

- Abrami, P.C., Leventhal, L. y Perry, R.P. (1982). Educational seduction. *Review of Educational Research*, 52, 444-464.
- Abrami, P.C. y D'Apollonia, S. (1990). The dimensionality of ratings and their use in personnel decisions. En M. Theall y J. Franklin (Comps.), *Student ratings of instruction: Issues for improving practice. New directions for teaching and learning*, 43 (pp. 97-111). San Francisco: Jossey-Bass.
- Aleamoni, L.M., Yimer, M. y Mahan, J.M. (1972). Teacher folklore and sensitivity of a course evaluation questionnaire. *Psychological Reports*, 31, 607-614.
- Aleamoni, L.M. y Hexner, P.Z. (1980). A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. *Instructional Science*, 9 (1), 67-84.
- Aparicio, J.J., Tejedor, F.J. y Sanmartín, R. (1982). *La enseñanza universitaria vista por sus alumnos: Un estudio para la evaluación de los cursos de la enseñanza superior*. Madrid: ICE/ Universidad Autónoma de Madrid.
- Apodaca, P. y Rodríguez, M. (1999). La opinión de los alumnos en la evaluación de la calidad docente: posibilidades, limitaciones y estructura dimensional de sus indicadores. En *Indicadores en la Universidad: información y decisiones* (pp. 311-327). Madrid: Consejo de Universidades/ Ministerio de Educación y Cultura.
- Barke, C.R., Tollefson, N. y Tracy, D.B. (1983). Relationship between course entry attitudes and end-of-course ratings. *Journal of Educational Psychology*, 75 (1), 75-85.
- Biglan, A. (1973). The Characteristics of Subject Matter in Different Academic Areas. *Journal of Applied Psychology*, 57 (3), 195-203.
- Braskamp, L.A. y Ory, J.C. (1985). *Evaluating teaching effectiveness: A practical guide*. Beverly Hills: Sage.
- Casero, A. (2008). Propuesta de un cuestionario de evaluación de la calidad docente universitaria consensuado entre alumnos y profesores. *Revista de Investigación Educativa*, 26, 25-44.
- Casero, A. (2010). ¿Cómo es el buen profesor universitario según el alumnado? *Revista Española de Pedagogía*, 246, 223-242.
- Cashin, W.E. y Perrin, B.M. (1983). Do college teachers who voluntarily have courses evaluated receive higher student ratings? *Journal of Educational Psychology*, 75 (4), 595-602.
- Castro, J.J. (1996). *Factores moduladores de la evaluación del profesor*. Tesis doctoral. Universidad de la Laguna, Tenerife.
- Centra, J.A. y Creech, F.R. (1976). *The Relationship Between Students, Teachers, and Course Characteristics and Student Ratings of Teacher Effectiveness*. Princeton, N.J.: Educational Testing Service.
- Cranton, P.A. y Smith, R.A. (1986). A new look at the effect of course characteristics on student ratings of instruction. *American Educational Research Journal*, 23 (1), 117-128.
- De Salvador, X. (1996). Sobre la evaluación de la actividad docente del profesorado universitario: ¿Está mediatizada la valoración de los alumnos por las calificaciones? *Revista Española de Pedagogía*, 203, 107-128.
- Díaz-Aguado, M.J. (1987). La percepción del profesor por el alumno: expectativas y actitudes. En J. Beltrán (Ed.), *Psicología Educativa: Vol. 1*. Madrid: UNED.

- Etopa, M.P. (2003). *Evaluación del profesorado de la Universidad de Las Palmas de Gran Canaria: Repercusiones del paso de una evaluación formativa a una evaluación sumativa*. Tesis doctoral. Universidad de Las Palmas de Gran Canaria, Islas Canarias.
- Feldman, K.A. (1978). Course characteristics and college students' ratings of their teachers and courses: What we know and what we don't. *Research in Higher Education*, 9, 199-242.
- Feldman, K.A. (1979). The significance of circumstances for college students' ratings of their teachers and courses: a review and analysis. *Research in Higher Education*, 10, 149-172.
- Feldman, K.A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education*, 18, 3-124.
- Feldman, K.A. (1986) The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: a review and synthesis. *Research in Higher Education*, 24, 129 - 213.
- Feldman, K.A. (1997). Identifying exemplary teachers and teaching: Evidence from student ratings. En R.P. Perry and J.C. Smart (eds.), *Effective Teaching in Higher Education: Research and Practice* (pp.368-395). New York: Agathon Press.
- Feldman, R.S. y Prohaska, T. (1979). The student as Pygmalion: Effect of student's expectancy on the teacher. *Journal of Educational Psychology*, 71, 485-493.
- Fernández, J. y Mateo, M.A. (1997). Student and faculty gender in ratings of university teaching quality. *Sex Roles*, 37 (11/12), 997-1003.
- García Valcarcel, A. (1989, junio). *Modelos didácticos en la Universidad de Cantabria* (Inf. de investigación). Universidad de Cantabria.
- Goebel, B.L. y Cashen, V.M. (1979). Age, sex and attractiveness as a factor in student rating of teachers. *Journal of Educational Psychology*, 71, 646-653.
- Gough, H.G. (1979). A creative personality scale for The Adjective Check List. *Journal of Personality and Social Psychology*, 37, 1398-1405.
- Hernández, P. (1991). *Psicología de la Educación: Corrientes Actuales y Teorías Aplicadas*. México: Trillas.
- Howard, G.S. y Bray, J.H. (1979). Use of norm groups to adjust student ratings of instruction: A warning. *Journal of Educational Psychology*, 71, 58-63.
- Marsh, H.W. (1980). The influence of student, course, and instructor characteristics in evaluation of university teaching. *American Educational Research Journal*, 17 (1), 219-237.
- Marsh, H.W. (1982). Factors affecting student's evaluations of the same course taught by the same instructor on different occasions. *American Educational Research Journal*, 19 (4), 485-497.
- Marsh, H.W. (1984). Student's evaluations of university teaching: dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.
- Marsh, H.W. (1987). Students' Evaluation of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research, *International Journal of Educational Research*, 11 (3), 255-388.
- Marsh, H.W. (2007). Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness. En R.P. Perry and J.C. Smart (eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, (pp.319-383). Netherlands: Springer.
- Marsh, H.W. y Overall, J.U. (1980). Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. *Journal of Educational Psychology*, 72, 468-475.
- Marsh, H.W. y Ware, J. (1982). Effects of expressiveness, content coverage and incentive on multidimensional student rating scales. *Journal of Educational Psychology*, 74, 107-116.

- Mateo, M.A. y Fernández, J. (1993). Dimensiones de la calidad de la enseñanza universitaria. *Psicothema*, 5 (2), 265-275.
- Naftulin, D.H., Ware, J.E. y Donnelly, F.A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, 48, 630-635.
- Neumann, L., y Neumann, Y. (1985). Determinants of students' instructional evaluation: A comparison of four levels of academic areas. *Journal of Educational Research*, 78 (3), 152-158.
- Pozo, C., Reboloso, E. y Fernández, B. (2000). The 'Ideal Teacher'. Implications for student evaluation of teacher effectiveness. *Assessment & Evaluation in Higher Education*, 25 (3), 253-263.
- Ryan, J.J., Anderson, J.A. y Birchler, A.B. (1980). Student evaluation: The faculty respond. *Research in Higher Education*, 12, 317-333.
- Salvador, L. (1990). *Los docentes universitarios exitosos desde la perspectiva del alumno: su caracterización psicopedagógica*. Tesis doctoral. Universidad de Salamanca.
- Theall, M. y Feldman, K.A. (2007). Commentary and update on Feldman's (1997) "Identifying exemplary teachers and teaching: Evidence from student ratings". En R.P. Perry and J.C. Smart (eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, (pp.130-143). Netherlands: Springer.
- Villa, A. (1985). La evaluación del profesor: perspectivas y resultados. *Revista de Educación*, 277, 55-93.
- Ware, J.E. y Williams, R.G. (1975). The Doctor Fox Effect: A study of lecturer effectiveness and ratings of instruction. *Journal of Medical Education*, 50, 149-156.
- Ware, J.E. y Williams, R.G. (1977). Discriminate analysis of student ratings as a means for identifying lecturers who differ in enthusiasm or information giving. *Educational and Psychological Measurement*, 37 (3), 627-639.
- Williams, R.G. y Ware, J.E. (1976). Validity of student ratings of instruction under different incentive conditions: a further study of the Dr. Fox effect. *Journal of Educational Psychology*, 68 (1), 48-56.
- Williams, R.G. y Ware, J.E. (1977). An extended visit with Doctor Fox: validity of student ratings after repeated exposures to a lecturer. *American Educational Research Journal*, 14 (4), 449-457.

ABOUT THE AUTHORS / SOBRE LOS AUTORES

Casero, Antonio (a.casero@uib.es). Licenciado en Psicología y Doctor en Ciencias de la Educación. Desde 1998 es profesor del Área de conocimiento de Didáctica y Organización Escolar en la Facultad de Educación de la *Universitat de les Illes Balears* (España). Miembro del grupo de investigación Educación y Ciudadanía (EIC). Su actividad investigadora se centra en los ámbitos de la calidad de la docencia universitaria, formación de formadores, ciberplagio, e inserción sociolaboral de colectivos en riesgo de exclusión social. Su dirección postal es: Departamento de Pedagogía Aplicada y Psicología de la Educación. Universitat de les Illes Balears. Crta de Valldemossa, km 7.5. 07122 Palma de Mallorca (Islas Baleares, España). [Buscar otros artículos de este autor](#)

[en Google Académico /Find other articles by this author in Scholar Google](#) 

ARTICLE RECORD / FICHA DEL ARTÍCULO

Reference / Referencia	Casero, Antonio (2010). Factores moduladores de la percepción de la calidad docente. <i>RELIEVE</i> , v. 16, n. 2. http://www.uv.es/RELIEVE/v16n2/RELIEVEv16n2_3.htm .
Title / Título	Factores moduladores de la percepción de la calidad docente. [<i>Modulating factors in the perception of teaching quality</i>].
Authors / Autores	Casero, Antonio
Review / Revista	RELIEVE (Revista ELecciónica de Investigación y EValuación Educativa / <i>E-Journal of Educational Research, Assessment and Evaluation</i>), v. 16, n. 1.
ISSN	1134-4032
Publication date / Fecha de publicación	2010 (Reception Date : 2009 September 25; Approval Date : 2010 September 30; Publication Date : 2010 September 30).
Abstract / Resumen	<p><i>This article examines the existing research findings in relation to the research of the factors that are presumably involved in university students' evaluations of the teaching quality they receive, possibly threatening the validity of the construct. The review is organized around the three implicated sources in the perceptive process of the student, being the student himself, the professor, and the context. The review concludes that, despite some minor effects, the factors under analysis do not represent a substantial threat to the system's validity. The unproductive debate that has gone on for decades about guarantees in the evaluation of teaching staff has largely been used to cover up the absence of a desire to institutionalize this kind of evaluation system.</i></p> <p>El presente artículo tiene como objetivo presentar una revisión sobre los hallazgos existentes en relación a la investigación de los factores supuestamente implicados en la valoración que el alumnado universitario realiza sobre la tarea docente de sus profesores, pudiendo suponer una amenaza a la validez de dicho constructo. La revisión se presenta organizada en torno a las tres fuentes implicadas en el proceso perceptivo del alumno, siendo éstas el propio alumno, el profesor y el contexto. La revisión concluye que, a pesar de efectos menores, los factores analizados no representan una amenaza substancial a la validez del sistema. El debate infructuoso que se ha producido durante décadas sobre las garantías en la valoración del personal docente ha sido usado en gran parte para encubrir la falta de deseo de institucionalizar este tipo de sistemas de evaluación.</p>
Keywords / Descriptores	<i>Student evaluation of Teacher Performance, Teaching Quality, Effective Teaching.</i> Evaluación del desempeño docente, Calidad docente, Efectividad docente
Institution / Institución	Departamento de Pedagogía Aplicada y Psicología de la Educación. Universitat de les Illes Balears (España).
Publication site / Dirección	http://www.uv.es/RELIEVE
Language / Idioma	Español (Title, abstract and keywords in English)

RELIEVE

Revista ELecciónica de Investigación y EValuación Educativa
E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).