

VALORACIÓN POR PARTE DEL PROFESORADO DE LA EVALUACIÓN DOCENTE REALIZADA POR LOS ALUMNOS

Juan Fernández*, Miguel A. Mateo* y José Muñiz**

* Universidad Complutense de Madrid, ** Universidad de Oviedo

Se estudió la valoración que los profesores universitarios hacen de la evaluación docente que llevan a cabo los estudiantes a través de las encuestas. El objetivo central fue someter a prueba la hipótesis general de la existencia de una actitud más favorable hacia los aspectos formativos de la evaluación que hacia los sumativos, así como comprobar si el modelo de evaluación condiciona dicha actitud. Un total de 2.651 profesores de las universidades Complutense de Madrid y de Oviedo respondieron a una escala de siete items tipo Likert que recogía sus opiniones sobre los distintos aspectos de la evaluación hecha por los alumnos. Los datos ofrecen un claro apoyo empírico a las hipótesis, tanto desde la perspectiva de la significación estadística como desde la consideración del tamaño del efecto. Se destacan las implicaciones de este tipo de metaevaluación para la mejora de la calidad docente.

Faculty assessment of students' teaching evaluation. An evaluation was carried out in which faculty members were asked to appraise students' university teaching evaluation. The aim was to test the hypothesis according to which teachers show a better attitude toward evaluation formative functions than to summative ones, and its corollary: This attitude will be affected by the evaluation model. 2.651 teachers from Complutense University of Madrid and from Oviedo University took part in this study. The data show empirical support for this hypothesis and its corollary, both in terms of statistical significance and effect size. This work emphasizes the implications of this kind of metaevaluation for improving teaching quality.

La evaluación de la calidad docente universitaria llevada a cabo a través de las opiniones de los alumnos es una realidad incipiente, aunque con visos de consolidación, en buena parte de las universidades españolas, como se anticipaba ya en el número monográfico de la revista *Studia Paedagogica* sobre evaluación de la enseñan-

za superior, publicado en 1988 (Fernández, 1988; Tejedor, Jato y Míguez, 1988).

Este tipo de evaluación tiene su origen formal en torno a los años 20 de este siglo en Estados Unidos, donde hoy es práctica ya bien asentada en la casi totalidad de sus universidades (Feldman, 1977), extendiéndose desde entonces hasta nuestros días por los cinco continentes (Angulo, Fernández y Martínez, 1987; Dawoud, 1983; Miller, 1988; Miron, 1988; Rushton y Murray, 1985; Watkins, Marsh y Young, 1987). Aunque, ciertamente, no es el úni-

Correspondencia: Juan Fernández
Facultad de Psicología
Universidad Complutense de Madrid
28223 Madrid. Spain

co sistema posible de evaluación de la docencia (Millman, 1981), sin embargo sí es el que ha recibido más apoyo empírico y el que goza en la actualidad de un mayor número de garantías en cuanto a la fiabilidad y validez de la información recogida (Marsh, 1987).

Sin embargo, esta forma de evaluar no está exenta de insuficiencias, una de las cuales es su marcado carácter unidireccional, dado que siempre son los alumnos los que evalúan la docencia impartida por sus profesores. Como hoy es ya bien sabido, el proceso de enseñanza/aprendizaje es lo suficientemente abarcador como para no poder ser evaluado en toda su complejidad por un único sistema básicamente unidireccional.

En un intento de superar esta clara deficiencia se ha comenzado a introducir en España la denominada *evaluación circular* (Fernández y Mateo, 1994), materializándose más específicamente en las universidades de Oviedo y Complutense de Madrid (Fernández, Mateo y Muñiz, 1995). Desde esta perspectiva se establece que todos los agentes que participan en el amplio contexto educativo tienen el derecho, y tal vez el deber, de evaluarse unos a otros recíprocamente, así como a las instituciones académicas, que a su vez tienen el cometido de llevar a cabo las pertinentes evaluaciones -docentes, investigadoras y de gestión- de dichos agentes (Fernández, 1992).

De esta forma se trata, por una parte, de abarcar en buena medida los distintos componentes significativos de cada proceso de enseñanza/aprendizaje, a la par que de cruzar las informaciones provenientes de los distintos agentes e instituciones, a fin de potenciar al máximo la riqueza interpretativa de los datos recogidos en el proceso evaluador.

Ahora bien, aunque parece bastante correcto afirmar que la así llamada evalua-

ción circular es más rica en matices que la evaluación de tipo unidireccional, todavía se echa en falta un ulterior proceso, conocido dentro del mundo de la evaluación como *metaevaluación* (valoración de la evaluación, ya sea ésta unidireccional o circular), a fin de garantizar al máximo las prescriptivas tomas de decisiones, consustanciales a cualquier forma de evaluación (Murray, 1984; Stufflebeam, 1978; The Joint Committee on Standards for Educational Evaluation, 1994).

Con este trabajo se trata justamente de iniciar la investigación empírica sobre el proceso metaevaluador con respecto al primer componente significativo de la evaluación circular -el referido a la valoración de la calidad docente- en aquellas universidades que ya se han iniciado en esta modalidad de la evaluación (Complutense y Oviedo).

Dentro del ámbito especializado del primer componente, es ya bien sabido que el aspecto *formativo* de la evaluación (es decir, que los datos recogidos se utilicen para mejorar lo evaluado) encuentra menos trabas y resistencias por parte del profesorado que el aspecto denominado *sumativo* (es decir, el referido a las tomas de decisiones administrativas en orden a paralizar, continuar o mejorar un contrato o, en el caso español, la concesión o negación de unas prestaciones económicas adicionales cada 5 o 6 años). De aquí surge la hipótesis a poner a prueba: los profesores universitarios valorarán más positivamente aquellos elementos del instrumento de medida elaborado *ad hoc* que tratan de operacionalizar la dimensión formativa que los que hacen referencia a la dimensión sumativa.

Un corolario de esta hipótesis es que la metaevaluación con respecto a ambos aspectos -formativo y sumativo- va a estar influida por el tipo de evaluación y de devolución de la información realizada.

Método

Sujetos

Un total de 2651 profesores tomaron parte en esta investigación, de los cuales 1964 pertenecían a la Universidad Complutense de Madrid y 687 a la Universidad de Oviedo. Desglosados en función de las distintas categorías académicas, fue posible contar con 340 catedráticos de universidad, 1245 titulares de universidad, 71 catedráticos de escuelas universitarias, 420 titulares de escuelas universitarias y 434 profesores no numerarios, tanto de facultad como de escuelas univesitarias. 141 profesores prefirieron no indicar su categoría, aunque sí cumplimentar el instrumento de evaluación, por lo que sus respuestas fueron incluidas en los correspondientes análisis de datos.

Instrumento

Dadas, por una parte, la especificidad de esta investigación y, por otra, la carencia de instrumentos idóneos dentro de la bibliografía especializada, se procedió a la elaboración de uno (Evaluación del Profesorado de la Valoración de los Alumnos - EPVA) que posibilitase la recogida de información de una manera aceptablemente válida y fiable. Este instrumento consta de 7 ítems, a los cuales se responde según una escala de valoración tipo Likert que va de 1 a 7, sabiendo que el 1 significa el total desacuerdo con lo expresado por cada elemento, mientras que con el 7 se manifiesta el total acuerdo.

El EPVA ha sido sometido a una serie de análisis a fin de poner de manifiesto sus propiedades métricas. Gracias a éstos, es posible afirmar, por un lado, que la consistencia interna resulta satisfactoria (el coeficiente Theta de Carmines alcanza un valor de 0.92) y, por otro, que los análisis de la dimensionalidad (PFA, rotación Varimax) parecen

apoyar la sostenibilidad de los dos núcleos teóricos referidos a las funciones formativa y sumativa, ya que dos de los tres factores obtenidos (Factor I -constituido por los elementos 1, 2 y 3-, y Factor II -definido por los elementos 6 y 7-) podrían recibir sin duda dichas denominaciones (ver Tabla 1).

Con respecto al Factor III hay que indicar que, si bien desde un punto de vista estrictamente conceptual los elementos que lo conforman (4 y 5) debieran aparecer englobados dentro del núcleo de la función formativa de la evaluación, sin embargo, la posible carga de deseabilidad social que comportan, obliga a una consideración independiente a la hora de la interpretación de los resulta-

Tabla 1
Estructura factorial de la escala EPVA

	Factores		
	I	II	III
1. Los resultados de la evaluación docente que se me entregan permite detectar los puntos fuertes y débiles de mi docencia80		
2. Gracias a la comunicación de los resultados de la evaluación es posible modificar los aspectos menos positivos de mi enseñanza90		
3. La evaluación me ha sido útil para realizar modificaciones en mi manera de impartir las clases78		
4. Resulta conveniente un servicio de ayuda profesional para cualquier profesor que quiera modificar su sistema de enseñanza.			.73
5. Estaría dispuesto a participar en actividades destinadas a la mejora de mi calidad docente77
6. Estimo que se debieran utilizar los resultados de la evaluación como parte a tener en cuenta en los procesos de promoción del profesorado85	
7. Debiera utilizarse este tipo de evaluación como criterio fundamental para la concesión de quinquenios de la evaluación docente84	

dos. Los tres factores alcanzan a dar cuenta del 99.72% de la varianza empírica.

Asimismo, los resultados de un análisis de conglomerados realizado concuerdan plenamente con lo mostrado por el análisis factorial. Siguiendo un procedimiento de formación en función de la mayor correlación (en valor absoluto) entre elementos, aparece un primer agrupamiento entre los elementos 6 y 7 ($r=0.83$), un segundo entre los elementos 1 y 2 ($r=0.80$), a los que se añade inmediatamente el 3 ($r=0.78$), y un tercero, sensiblemente distante de los anteriores, entre los elementos 4 y 5 ($r=0.58$).

Procedimiento

A todos los profesores cuya labor docente estaba siendo evaluada por sus alumnos durante el curso académico 1994/95, se les dio la oportunidad de valorar a su vez la evaluación realizada por los estudiantes. De esta forma, dentro de un único proceso evaluador, cada profesor evaluado podía cumplimentar la encuesta, si así lo deseaba, entregándola ulteriormente a los mismos aplicadores que recogían los cuestionarios que cumplimentaban los alumnos. La recogida de la información, que tuvo lugar en fechas próximas al examen final de cada asignatura, se realizó en el aula normal de clase, en tiempo todavía lectivo.

Resultados

En primer lugar, y en referencia a la hipótesis considerada, se compararon las valoraciones (promedio) que merecen para los profesores aquellos elementos del EP-VA referidos estrictamente a los aspectos formativos de la evaluación (elementos 1, 2 y 3) con las dadas a los elementos relativos a los aspectos sumativos (elementos 6 y 7). Los resultados son ciertamente ilustrativos. La valoración media del núcleo formativo (4.50) es mayor que la del núcleo sumativo

(4.04), siendo la diferencia estadísticamente significativa ($t=13.56$, $p=0.0000$), a lo que hay que añadir un tamaño del efecto ($d=0.44$) que, de acuerdo con Cohen (1988), puede ser interpretado como medio.

En segundo lugar, y en relación con el corolario propuesto, se compararon entre sí las valoraciones obtenidas en las dos universidades (Complutense y Oviedo), ya que los respectivos contextos evaluativos generales son marcadamente distintos. Teniendo en cuenta simultáneamente todos los elementos del instrumento de evaluación es posible observar la existencia de diferencias multivariantes estadísticamente significativas ($T^2=132.15$, $F=26.39$, $p=0.0000$). Este resultado global fue desglosado, considerándose por separado las valoraciones (promedio) de los dos núcleos teóricos. Así, para el núcleo formativo aparecieron diferencias estadísticamente significativas ($t=10.35$, $p=0.0000$) entre Complutense (4.69) y Oviedo (3.95), con un tamaño del efecto medio ($d=0.46$), al igual que para el núcleo sumativo (Complutense=4.18, Oviedo=3.66, $t=5.75$, $p=0.0000$), aunque el tamaño del efecto es algo más moderado en esta ocasión ($d=0.26$).

Un desglose aún más detallado, por ítems, muestra resultados que señalan la misma dirección (excepto para los ítems 4 y 5 -ver Tabla 2).

Ítems	Complut.	Oviedo	t	p	d
1	4.70	3.95	9.81	0.0000	0.44
2	4.80	4.13	8.57	0.0000	0.38
3	4.57	3.76	10.02	0.0000	0.45
4	4.80	4.70	1.07	0.2831	0.05
5	5.49	5.45	0.47	0.6360	0.02
6	4.40	3.89	5.30	0.0000	0.24
7	3.96	3.44	5.42	0.0000	0.24

En resumen, cabe afirmar que se han encontrado en el presente trabajo indicios razonables de que tanto la hipótesis central propuesta como su corolario reciben apoyo empírico, es decir, los profesores universitarios tienden a valorar más favorablemente los aspectos formativos de la evaluación de su calidad docente que aquellos otros de índole sumativa, y esta tendencia se ve modulada por las características particulares del contexto en que aquélla se lleva a cabo.

Análisis de los Resultados

En cuanto al nuevo instrumento de evaluación - el EPVA- cabe indicar, a la luz de los datos referidos a sus propiedades formales, que posibilita recoger de forma relativamente válida y fiable la información pertinente en torno a los aspectos formativos y sumativos de la evaluación, lo cual cobra una considerable importancia cuando se constata, tras la pertinente consulta bibliográfica nacional e internacional, la escasez de este tipo de instrumentos de medida.

Por lo que atañe a la hipótesis central de este trabajo, es posible afirmar que los resultados conjuntos de la universidades Complutense y de Oviedo concuerdan con los que aparecen reflejados en la bibliografía especializada. Los profesores valoran de forma más positiva las funciones formativas de la evaluación que las sumativas. Parece existir un consenso nacional e internacional a la hora de entender que los datos de la evaluación de la calidad docente, llevada a cabo por medio de las opiniones de los alumnos, pueden resultar útiles para el profesor, ya que a través de los mismos, es posible detectar los *puntos fuertes y débiles* de la enseñanza impartida, y con ello, inferir las vías para realizar avances en el perfeccionamiento docente. La utilización de los datos de la evalua-

ción para fines sumativos, en el caso concreto español para la obtención de quinuenios (una remuneración adicional cada 5 años) y para la promoción del profesorado (oposiciones), muestra una valoración menos positiva. Aparecen, ciertamente, claras diferencias estadísticamente significativas entre ambas valoraciones, aunque la consideración de las puntuaciones desglosadas de la tabla 2 permite constatar que los distintos valores encuadrables dentro del factor sumativo rondan, en general, el valor de la media teórica de la escala de valoración.

Ahora bien, como se había previamente anticipado, esta tendencia general a valorar más positivamente los aspectos formativos que los sumativos de la evaluación muestra diferencias considerables, estadísticamente significativas, entre los dos subgrupos de sujetos considerados (universidades Complutense y de Oviedo), en los cinco elementos que componen los dos factores.

En todos los casos la valoración más positiva es la que presenta el grupo de la Complutense, como es posible observar en la tabla 2. Dado que las diferencias en el modelo de la devolución de la información son considerables entre ambas universidades, parece lógico inferir que sea ésta la variable determinante de esas diferencias. Estos datos parecen ofrecer un considerable apoyo a los especialistas evaluadores que atestiguan que el formato de la devolución de la información se constituye en una condición necesaria, aunque en modo alguno suficiente, para la ulterior utilización de la evaluación por parte de sus receptores.

Especial consideración merecen los resultados referidos a los dos elementos del Factor III del EPVA. En primer lugar, cabe afirmar que, en ambos grupos, las valoraciones medias de los elementos 4 y 5 sobrepasan con creces el punto de corte

de la positividad/negatividad valorativa (que es 4). De hecho, son los valores más altos de todos los encontrados, si se exceptúa la igualdad entre las puntuaciones de los elementos 2 y 4 en el grupo de la Complutense. En segundo lugar, es necesario resaltar que en estos dos elementos es donde se muestra la mayor semejanza valorativa entre los profesores de ambas universidades. Es conveniente destacar aquí que estos datos apoyan de nuevo uno de los principios ya clásicos dentro del

ámbito de la evaluación que afirma que no es fácil obtener mejoras en la calidad docente mediante la sola devolución de la información proporcionada por los alumnos. La experiencia una y otra vez ha demostrado que es necesario un servicio de ayuda profesional si se quiere optimizar la docencia, tal cual es solicitado en este trabajo por una abrumadora mayoría del profesorado tanto de la universidad de Oviedo como de la universidad Complutense.

Referencias

- Angulo, F., Fernández, J. y Martínez, M.R. (Eds.). (1987). *La evaluación de la enseñanza universitaria*. Madrid: Instituto de ciencias de la Educación.
- Dawoud, M.(1983). Une grille d'évaluation de la compétence du professeur d'université. *Revue de Psychologie Appliquée*,33,17-31.
- Feldman,K.A.(1977). Consistency and variability among college students in rating their teachers and courses: A review and analysis. *Research in Higher Education*,6,223-274.
- Fernández,J.(1988). La evaluación de la enseñanza universitaria: la experiencia de la Complutense. *Studia Paedagogica*,20,135-146.
- Fernández,J.(1992). Evaluation and decision making in the Complutense University of Madrid. *Higher Education Management*, 4,336-345.
- Fernández,J. y Mateo,M.A.(1994). The circular evaluative process of higher education: A Spanish sample. In OECD(Ed.), *Evaluation and decision making process in higher education: French, German and Spanish experiences*. Paris: OECD.
- Fernández,J., Mateo,M.A. y Muñoz,J.(1995). Evaluation of the academic setting in Spain. *European Journal of Psychological Assessment*,11,134-138.
- Joint Committee on Standards for Educational Evaluation (1994). *The program evaluation standards -How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks,CA: SAGE.
- Marsh,H.W.(1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*,11,253-288.
- Miller,A.H.(1988). Student assessment of teaching in higher education. *Higher Education*,17,1-13.
- Millman,J.(Ed.).(1981). *Handbook of teacher evaluation*. Beverly Hills,CA: SAGE.
- Miron,M.(1988). Students' evaluation and instructors' sel-evaluation of university instruction. *Higher Education*,17,175-181.
- Murray,H.G.(1984). The impact of formative and summative evaluation of teaching in North American universities. *Assessment and evaluation in Higher Education*, 9,117-132.
- Rushton,J.P. y Murray,H.G.(1985). On the assessment of teaching effectiveness in British universities. *Bulletin of the British Psychological Society*,38,361-365.
- Stufflebeam,D.L.(1978). Meta-evaluation: An overview. *Evaluation and the Health Professions*,1,17-43.
- Tejedor,F.J., Jato,E. y Míguez,C.(1988). Evaluación del profesorado universitario por los alumnos en la universidad de Santiago. *Studia Paedagogica*,20,73-134.
- Watkins,D.,Marsh,H.W. y Young,D.(1987). Evaluating tertiary teaching: A New Zealand perspective. *Teaching & Teacher Education*,3,41-53.