

Evaluación del profesorado: posible fórmula correctora

.....
José Juan Castro Sánchez

UNIVERSIDAD DE LA LAGUNA

RESUMEN

Este artículo pone a prueba una posible fórmula correctora de los resultados obtenidos por tres profesores tras ser evaluados por sus alumnos. Esta fórmula correctora está basada en la distancia que la evaluación de cada alumno mantiene respecto a la evaluación media del grupo dividida por la desviación típica del grupo, es decir, las conocidas como puntuaciones Z.

ABSTRACT

This article puts on approval a possible formula proofreader of the results obtained by three professors after being evaluated by its students. This formula proofreader is based on the distance that each student's evaluation maintains regarding the half evaluation of the group divided by the typical deviation of the group, that is to say, the acquaintances as punctuations Z.

Introducción

El papel de los alumnos en la evaluación de la calidad de la educación superior es múltiple y variado. A veces es propuesto como mero elemento pasivo que más que evaluar es evaluado, presuponiendo que su mayor o menor calidad en una serie de parámetros, bien anteriores a su acceso a la educación universitaria, bien durante su paso por la universidad o bien tras la conclusión de los estudio, es un indicador de la menor o mayor calidad de la educación superior, en general, o de una universidad, en particular.

Otras veces es propuesto como un elemento sobre el que medir todo lo que una determinada institución pone a su servicio para su mejor formación: planes de estudio, profesorado, servicios complementarios, actividades extraescolares, etc... Bajo este modelo se presupone que el trato que recibe el alumno, en general, es una buen indicador, junto a otros, de la calidad de una institución universitaria. Aunque en esta evaluación el alumno, que parece ser el receptor principal de la calidad que la institución ofrece, puede tener un papel más activo, predominan los modelos donde se mide la calidad de todo lo puesto a su disposición, pero sin contar con la opinión de estos alumnos.

Por último, y bastante más frecuente en las universidades de nuestro país, y más en concreto, en las de nuestra comunidad, el alumno como un activo evaluador de elementos tan variados como: satisfacción con la carrera, metodología docente, profesorado, en general y en particular, elementos contextuales, causas de la mala calidad, causas del fracaso, etc.. Aunque en este último enfoque el papel del alumno parece sumamente activo y participativo, probablemente se haya llegado a él por lo económico de su empleo, generalmente a través de encuestas de opinión, y el futuro más probable es que cada vez se vaya reduciendo su papel, hasta ahora casi el único evaluador, en la medida en que se incorporen otros evaluadores: profesorado de la propia universidad, colegas de otras universidades, expertos externos, representantes de la sociedad y del mundo empresarial, etc...

Es este último papel el más directamente relacionado con nuestra investigación por lo que lo abordaremos de una manera más profunda. En este sentido, diferentes estudios realizados en las universidades de Barcelona y Oviedo (Rodríguez, 1989; De Miguel y Rodríguez, 1990) proponen una evaluación ponderada de las componentes docente, investigadora y de servicios del profesorado. Con estas propuestas podemos hacernos una idea del posible peso del alumnado como evaluador. Según estos autores:

- 1) 50% para la docencia, evaluada por:
 - Estudiantes, 50%.
 - Colegas, 30%.
 - Uno mismo, 20%.
- 2) 40% para la investigación, evaluada por:
 - Expertos externos, 50%.
 - Colegas, 25% .
 - Uno mismo, 25% .
- 3) 10% para servicios públicos y administrativos valorados por colegas y/o expertos.

En la Universidad de Barcelona, por ejemplo (Rodríguez, 1989), el profesorado acepta sin problemas la evaluación formativa y una amplia mayoría (dos tercios aproximadamente) no rechaza la función sumativa y de control. El 60% acepta el que los alumnos (incluso al profesor) tengan carácter público. Se cree que las autoridades académicas, seguido de los colegas del departamento, son la mejor audiencia para los resultados de la evaluación. Por último, se acepta a los alumnos como los mejores evaluadores de la docencia.

Otro estudio es el realizado en la Universidad de Valencia (Jornet y otros, 1989). Algunos resultados destacados referentes al peso evaluador del alumno frente a otros evaluadores son los siguientes: Que los alumnos intervengan en la evaluación docente es aceptado por un 67% de los encuestados. Como manifestación del clima poco cordial que existe dentro de los departamentos universitarios, un 45% se manifiesta en contra de la idea de ser evaluados por los colegas del propio departamento. La evaluación por comisiones de colegas democráticamente elegidos es aceptada por un 48% del profesorado, aunque un 42% de los catedráticos se oponen a ella. La autoevaluación es aceptada por el 58%.

En la Universidad Politécnica de Valencia (Fernández y otros, 1991), se ha encontrado más reticencia entre el profesorado hacia otras fuentes de evaluación de la docencia, que a los propios alumnos.

Escudero (1993) cita a los alumnos como la primera y principal fuente de información en la evaluación de la enseñanza. Y a continuación opina que los alumnos son los mejores jueces de aspectos como los métodos de enseñanza, la rectitud y equidad pedagógica, el interés del profesorado en los alumnos, el interés que despierta la asignatura, la valoración global del curso, etc...

En un estudio realizado en 1979 por Marques, Lane y Dorfman (Díaz-Aguado, 1987) con una muestra de 40 alumnos y 40 profesores de distintas facultades en la que les pidió a los sujetos que calificaran 100 perfiles de profesores hipotéticos, que contenían 7 puntuaciones en dimensiones relacionadas con la eficacia docente, se encontró un gran acuerdo en los criterios utilizados para evaluar la eficacia por alumnos y profesores.

El estudio realizado por Marques, Lane y Dorfman permite concluir que existe un gran acuerdo en los criterios utilizados por alumnos y profesores al juzgar la eficacia de profesores hipotéticos; y según los resultados obtenidos en otros estudios (Marsh, Overall y Kesler, 1979) parece que el acuerdo se mantiene incluso entre el propio profesor y sus alumnos a la hora de evaluar la eficacia del primero; así como también entre los alumnos y otros profesores al juzgar la eficacia de profesores reales (Marsh, 1982).

Como vemos, los alumnos son ampliamente aceptados como activos agentes evaluadores de la calidad de la educación. Sin embargo, las discrepancias aparecen cuando se trata de calificar a los alumnos como jueces eficaces y no sesgados. Generalmente, se ha dudado del alumnado como fuente de evaluación veraz y consistente, a pesar de la numerosa literatura en contra.

Good y Mulryan (1990) señalan que el interés por el estudio de la efectividad de los profesores, vista por lo alumnos, empieza en el siglo pasado y que en 1899 Kratz publicó un trabajo sobre el tema. Posteriormente se empezaron a elaborar escalas: Elliot en 1915, Rugg en 1920, Barr en 1931. Según estos autores, fue particularmente famosa la de Boyce en 1915 (Boyce Card), que tenía varias secciones: características personales, sociales y profesionales, gestión escolar, técnicas de enseñanza y resultados; y 44 aspectos o cuestiones de evaluación.

Cook (1989) señala, no obstante, que la primera escala verdaderamente diseñada para estudiantes fue la "Pordue Rating Scale of Instruction", publicada en 1926.

A partir de aquellos años, la extensión y desarrollo del procedimiento ha sido enorme, de forma que hoy son mayoría las instituciones de educación superior del mundo occidental que usan, al menos con cierta regularidad, encuestas a los estudiantes para evaluar la docencia y/o sus profesores (Abram 1989) y son muchos miles de páginas de literatura especializada las que recogen cada año los trabajos de investigación sobre el tema.

Hasta el momento son varias las revisiones generales de la investigación sobre distintos aspectos relacionados con la utilización de encuestas a los alumnos, con lo que existe un conocimiento bastante cabal del estado global de la cuestión.

En 1981, Cohen presentó un meta-análisis de 41 estudios sobre la relación entre las evaluaciones de los alumnos y su aprendizaje, encontrando una correlación promedio de 47.

Los trabajos de Cohen fueron replicados y extendidos por Feldman (1986, 1989a y 1989b), con 19 estudios en la primera fase y 28 en la última versión. Estudió las relaciones entre las evaluaciones de seis fuentes diferentes: los estudiantes actuales, los antiguos estudiantes, el propio profesor, los colegas, los administradores y los observadores externos. Uno de los resultados más importantes fue la moderada relación ($r = .29$ de promedio) entre las auto-evaluaciones y los juicios de los estudiantes.

Otro de los investigadores que más han venido trabajando en este terreno ha sido H.W. Marsh, autor de varios análisis de síntesis y revisiones de la investigación y autor (Student Evaluation of Quality), que ha sido realizado "cros culturalmente" incluso en nuestro país (Marsh y otros, 1985; Touron, 1989).

En una de sus revisiones, Marsh (1987) habla de las buenas características de las encuestas de los estudiantes para valorar la docencia, sobre todo si se comparan con los juicios de otros colectivos, pero no oculta las dificultades de validación de las valoraciones de los estudiantes, debido a la inexistencia de un criterio universal de "calidad docente".

Como sugerencia propone que los juicios de los alumnos sean complementados por evaluadores externos.

Las principales relaciones encontradas por Marsh entre las valoraciones de los alumnos y distintas características básicas fueron las siguiente:

- Con respecto al interés previo por la materia, encontró que las clases con más interés por la materia evalúan de manera más favorable, pero no está claro si el interés existía antes o es generado tras cursar la asignatura.

- Con respecto a la calificación esperada/real, afirma que las clases con calificaciones más altas evalúan algo mejor, pero no sabe bien si por "indulgencia" o por superior aprendizaje.
- Con respecto a la razón para cursar la asignatura, observó que, las asignaturas más selectivas y las más pretendidas, en general, tienden a ser valoradas algo mejor.
- Con respecto al trabajo/dificultad, halló que las materias más duras y más difíciles tienden a ser valoradas algo mejor.
- Con respecto al tamaño de la clase, aparecen hallazgos contrapuestos, predominan los estudios que encuentran valoraciones más positivas en las clases más pequeñas, pero otros encuentran una relación curvilínea entre tamaño y valoración, que favorece también a las clases más numerosas.
- Con respecto al rango del profesor, encuentra efectos muy débiles y contrapuestos.
- Con respecto al sexo del profesor y/o del alumno, de nuevo encuentra efectos muy débiles y hallazgos contrapuestos.
- Con respecto a las áreas de estudio, hay una ligera tendencia a valoraciones más altas en humanidades y más bajas en ciencias, aunque hay muy pocos estudios realizados.
- Con respecto a los propósitos de la evaluación, se sabe que cuando las valoraciones van a servir para estabilidad y/o promoción del profesor se valora algo más alto.
- Con respecto a las condiciones de aplicación, se evalúa de manera más alta cuando los cuestionarios no son anónimos y cuando el profesor/a está presente.
- Por último, con respecto a la personalidad del estudiante, halló efectos muy débiles y hallazgos contrapuestos.

Cook (1989) se refiere también al posible problema de indulgencia en los estudiantes, así como al efecto halo, esto es, a la tendencia a valorar todo bajo una impresión global, en vez de distinguir entre cada aspecto. Sin embargo, los estudios realizados parecen apoyar más la hipótesis contraria (Díaz Aguado, 1987)

Además de los estudios citados, existen otros (Murray, 1984; Wigington y otros, 1989) que destacan la gran utilidad y las buenas cualidades de los juicios de los estudiantes para evaluar la docencia universitaria.

Kurtz, Mueller, Gibbons y Dieatardo (1989), han puesto de manifiesto que las evaluaciones de los alumnos presentan un nivel significativo de consistencia interna de las escalas, entre los alumnos evaluadores, diferentes períodos del curso y distintos cursos con tipologías similares.

De igual manera, se ha encontrado que las evaluaciones de los alumnos son bastante estables al comparar evaluaciones realizadas con varios años de diferencia (Overal y Marsh, 1980) e incluso después de una experiencia tan importante como la del ejercicio de la propia carrera (Firth, 1979).

Loder (1990) señala que los alumnos son los mejores jueces para los aspectos tales como los métodos de enseñanza, la rectitud y equidad pedagógica, el interés del profesorado en los alumnos, el interés que despierta la asignatura, la valoración global del curso, etc...

No obstante, y a pesar de las revisiones anteriores, algunos autores como Cahn (1987) no aceptan, de ninguna forma, la evaluación de los alumnos. En el lado opuesto, Marsh (1987), Arubayi (1987) y Franke-Wikberg (1990) defienden permanentemente la consistencia, estabilidad y validez de las evaluaciones realizadas por los alumnos.

La influencia de las calificaciones obtenidas en la evaluación del profesor por los alumnos ha sido, probablemente, la variable más estudiada. Una gran parte de los estudios realizados hasta el momento encuentran correlaciones positivas entre las calificaciones obtenidas por los alumnos, o que éstos esperan obtener, y la evaluación del profesor (Kelley, 1972). Esta correlación puede explicarse de formas muy diferentes:

- a) Los alumnos evalúan favorablemente al profesor porque han sido favorablemente evaluados por él, es decir, con benevolencia.
- b) El profesor evalúa con benevolencia a los alumnos porque ha sido, o espera ser, benevolamente evaluado por ellos.
- c) Los alumnos con mejores notas evalúan mejor a su profesor porque han aprendido más con él. En este último caso, la correlación no sería una prueba de invalidez.

Por otra parte, también conviene considerar que no todos los estudios han encontrado correlación positiva entre las calificaciones de los alumnos y la evaluación que éstos hacen del profesor (Díaz Aguado, 1987).

Holmes (1972) propone no estudiar la relación directa entre calificaciones obtenidas y evaluación del profesor por los alumnos, sino calificación esperada versus calificación obtenida y evaluación del profesor.

En este sentido, Díaz-Aguado (1986) ha podido comprobar que el rendimiento de los alumnos, en contra de lo que cabía esperar, apenas influye en la evaluación que se hace del profesor. Lo que sí influye es la diferencia entre la nota obtenida y la esperada. Si no se ha obtenido la nota esperada, se juzga de manera más dura al profesor.

Con respecto a la motivación, se sabe que afecta a la evaluación del profesor. Los alumnos más motivados, no es que sean evaluadores más benévolo, sino más discriminativos en todos los sentidos y en todas las direcciones (Díaz-Aguado, 1986).

También se sabe que las clases poco numerosas evalúan de manera más favorable que las clases más numerosas. Sin embargo, es difícil interpretar esta correlación, puesto que el número de alumnos de la clase correlaciona negativamente con otras medidas de eficacia docente (Díaz-Aguado, 1987).

La evaluación del profesor por el alumno se ve favorecida por la similitud entre el profesor y el alumno en variables tales como el estilo cognitivo o las actitudes hacia la educación (Hoffman y Kremer, 1980)

Por último, el momento en que se realiza la evaluación (Carrier, Howard y Miller, 1974): la persona que aparece como responsable de la evaluación (Abrami, Perry y Breen, 1976) y el anonimato de las respuestas (Stone, Rabinowitz y Spool, 1977) parecen influir sobre la evaluación del profesor, por lo que son variables que conviene controlar a la hora de comparar distintas evaluaciones (Díaz-Aguado, 1987).

Hernández (1986) señala que en investigaciones no universitarias ha encontrado lo que todas las universidades hallan, que es la gran fiabilidad en las respuestas de los alumnos, por encima de .80. Esto quiere decir, concluye, que los alumnos entre sí tienen una gran consistencia en la evaluación. En otras palabras, hay poca desviación típica, están todos muy concentrados en una misma puntuación.

Báez y otros (1996), entre las conclusiones de su propuesta de actuación para la evaluación y mejora de la calidad de la docencia de la Universidad de La Laguna concluyen: los estudiantes de esta Universidad pueden ser considerados jueces fiables y no sesgados de la actuación docente del profesorado. Esta conclusión viene avalada por los resultados obtenidos en los diversos análisis a que fue sometida la escala de desempeño docente: consistencia interna alta, discriminación adecuada del profesorado (según curso, división de estudios, antigüedad, grado académico y categoría administrativa) y estrecha relación con la opinión de los propios profesores. Un dato adicional que apoya esta conclusión es que el nivel de exigencia o la dificultad en la materia no se asocia de forma sistemática ni al buen ni al mal desempeño docente.

En definitiva, la investigación sobre los posibles sesgos evaluativos del estudiante demuestra que las opiniones de alumnado pueden considerarse razonablemente estables, fiables y acreditadas (Marsh y Roche, 1992; Smith y Cranton, 1992). En este sentido, las evaluaciones de los estudiantes pueden ser tan sistemáticas como las de otros colegas docentes.

A modo de conclusión, Ramsden (1991) dice que existen serios argumentos a favor de la evaluación por alumnos y los concreta en cuatro:

- a) El alumno tiene capacidad para opinar sobre si la docencia recibida le ha ayudado o no en el proceso de aprendizaje.
- b) El alumno distingue entre la adecuada actuación del profesor y la docencia efectiva.
- c) La investigación pone de manifiesto un alto nivel de consistencia entre los resultados de la evaluación de la docencia y los rendimientos de los alumnos
- d) Sin embargo, se considera impropio preguntar a los alumnos sobre cuestiones docentes para las cuales no están capacitados para opinar.

Respecto a la validez, los trabajos de Cohen (1981), Ramsden y Entwistle (1981), Entwistle y Ramsden (1983) y Entwistle y Tait (1990), entre otros, ponen de manifiesto el nivel de validez moderadamente significativo de las evaluaciones de los alumnos como medida de la efectividad de los profesores. En sentido contrario se sitúan Dowell y Neal (1982), entre otros.

Baxter (1991) dice que la cuestión de la validez de las evaluaciones de los alumnos ha de estar integrada en un proyecto donde se especifique la finalidad de los resultados, considerándola como un aspecto que ha de ser complementado con otras valoraciones.

Ante el crítico asunto de la validez existen básicamente dos visiones (Abrami y otros, 1990). La primera indica que las evaluaciones de los alumnos serán válidas si reflejan con precisión sus opiniones sobre la calidad de la enseñanza, con independencia de que el juicio refleje o no su aprendizaje. La segunda indica que las evaluaciones serán válidas si reflejan de manera precisa la eficacia instruccional.

Con respecto a esta polémica, Escudero (1993) concluye que las encuestas a alumnos reflejan bien lo que ven y viven en el aula y, de hecho, los juicios de los alumnos son más sensibles a los cambios de enseñanza que su propio rendimiento, por lo que entienden que las críticas a la validez se hacen si se asume la segunda visión, pero tales críticas no se dan si se asume la primera visión.

No obstante conviene no generalizar, no todos los alumnos son igual de fiables y no sesgados. Hay circunstancias muy particulares y específicas de alumnos concretos que probablemente tiñan la evaluación de los alumnos. Probablemente estas circunstancias queden, en la mayoría de los casos, contrarrestadas entre sí, no afectando de manera significativa el resultado medio final. Es decir, frente a alumnos que infravaloran hay otros que supervaloran y el resultado medio final no nota, en la mayoría de los casos, estos sesgos.

Sin embargo se sabe que las evaluaciones del profesor, a pesar de la consistencia evidenciada en la mayoría de los estudios, queda, a veces, explicada, en parte, por factores extraños al propio ejercicio del profesor. Hernández (1986) encontró, en niveles no universitarios, que las características de inadaptación de algunos alumnos explican hasta el 80% de la varianza de la evaluación del profesor.

Por tanto, podemos concluir que el alumnado como grupo es un elemento evaluador fundamental en todo sistema que pretenda evaluar y mejorar la calidad de la enseñanza universitaria. La evaluación del profesor por sus alumnos parece ser una importante fuente de información para juzgar determinados aspectos de la eficacia docente: como, por ejemplo, la activación de los intereses de los alumnos, su estilo de interacción y su forma de dar la clase. La estabilidad y precisión que parecen tener las evaluaciones hacen aconsejable su utilización en una aproximación multidimensional de la calidad que incluya también otras medidas.

Sin embargo, no debemos olvidar que probablemente algunas circunstancias muy particulares pueden influir en las evaluaciones que algunos alumnos concretos hacen. Citando de nuevo a Marsh (1987), parte de las puntuaciones de los alumnos pueden estar condicionadas, poseer cierta falta de fiabilidad y estar afectados por algunas fuentes potenciales de sesgos.

De ahí que sea muy útil la búsqueda de fórmulas que permitan aislar estos minoritarios juicios sesgados y separarlos de los mayoritariamente no sesgados para obtener una evaluación más precisa y objetiva de lo evaluado.

El objetivo central de nuestra investigación es encontrar un mecanismo, que nos permita, cuando los alumnos evalúan al profesor, separar los minoritarios juicios sesgados de los mayoritariamente no sesgados, de tal manera que junto a la valoración de toda el aula, podamos ofrecer al profesor evaluado una valoración modulada (más objetiva, calculada) con solo los alumnos no sesgados.

Para encontrar este mecanismo analizamos un proceso de evaluación real. Es importante señalar la independencia de los autores de la evaluación real llevada a cabo. La independencia en la construcción de los cuestionarios en la aplicación de los posterior recogida de datos permitirá extrapolar las conclusiones a cualquier otra evaluación, puesto que la pretensión es que los hallazgos sean independientes de los cuestionarios aplicados, en definitiva, del modelo de evaluación puesto en práctica.

Metodología

Sujetos

80 alumnos universitarios que cursaban una misma asignatura.

Instrumento

El cuestionario fue construido por los profesores evaluados, en él los alumnos expresaban su opinión acerca de: conocimientos y competencia; motivación y entusiasmo del profesor; motivación que despierta en los alumnos; organización y estructuración de la exposición; interacción profesor-alumno en clase; y relación afectiva profesor-alumno. El cuestionario está planteado en escala tipo Likert, de 1 a 5. Donde 1 significaba muy en desacuerdo, 2 en desacuerdo, 3 ni en desacuerdo ni de acuerdo, 4 de acuerdo y 5 muy de acuerdo. El cuestionario evaluaba seis aspectos de la actividad docente.

Procedimiento

Los alumnos debían puntuar de 1 a 5 a cada uno de los tres profesores que le habían impartido la asignatura en las seis dimensiones evaluadas.

Análisis estadístico

Se aplicó el programa estadístico SPSSWIN96,

Resultados

Los resultados figuran en la Tabla N° 1. En ella podemos observar como las puntuaciones son, en general, positivas. En todos los casos, las valoraciones medias de cada profesor en cada dimensión superan el valor intermedio de 3.

TABLA N° 1

	PROFESOR 1	PROFESOR 2	PROFESOR 3
Conocimientos y competencia	4.8125	4.5625	3.9750
Motivación y entusiasmo del profesor	4.5750	4.6625	4.0500
Motivación que despierta en los alumnos	4.2625	4.2250	3.6250
Organización y estructuración de la exposición	4.1750	3.9375	3.2625
Interacción profesor-alumno en clase	4.2000	4.1875	4.0000
Relación afectiva profesor-alumno.	3.4250	4.0125	3.7750
VALORACION GLOBAL	4.2417	4.2646	3.7813

A partir de aquí iniciamos el tratamiento estadístico dirigido a depurar las anteriores puntuaciones iniciales. Para ello, convertimos la valoración inicial de cada alumno en una puntuación típica o puntuación Z en función de la media y la desviación típica obtenida por cada profesor en cada dimensión.

Luego, tomamos como primer criterio de realidad evaluativa a aquellos alumnos que en la mitad de las dimensiones evaluadas y en la mitad de los profesores evaluados emitían juicios valorativos ajustados a la media, es decir, sin sobrepasar las puntuaciones Z-1 y +1. Es decir, un mínimo de calidad evaluativa del 50%.

Tras esta transformación, 67 alumnos superaron este requisito. La valoración que hicieron estos alumnos de los tres profesores figura en la Tabla N° 2.

De nuevo, las valoraciones en general son de signo positivo. Los cambios habidos nos superan 3% con respecto a las valoraciones iniciales. El profesor 1 mejora en 4 de las seis dimensiones y en la valoración global. Los profesores 2 y 3 empeoran su valoración en 5 de las seis dimensiones y en la valoración global.

El siguiente criterio de calidad fue más exigente. Los alumnos debían evaluar de manera ajustada a la media, es decir, no sobrepasar el intervalo de puntuación Z-1 +1, en dos o más profesores y en 4 o más dimensiones. Es decir, un mínimo de calidad evaluativa del 66.7%.

Tras esta selección, 39 alumnos (aproximadamente el 50%) superaron tal criterio de calidad. En la Tabla N° 3 figuran los resultados de los profesores al ser evaluados por estos alumnos.

TABLA N° 2

	PROFESOR 1	PROFESOR 2	PROFESOR 3
Conocimientos y competencia	4.8360	4.5520	4.5640
Motivación y entusiasmo del profesor	4.5970	4.6720	4.6410
Motivación que despierta en los alumnos	4.2540	4.1719	4.1030
Organización y estructuración de la exposición	4.2540	3.9250	3.9490
Interacción profesor-alumno en clase	4.1940	4.1490	3.9230
Relación afectiva profesor-alumno.	3.5220	3.9850	3.7180
VALORACION GLOBAL	4.2760	4.2440	4.1500

TABLA N° 3

	PROFESOR 1	PROFESOR 2	PROFESOR 3
Conocimientos y competencia	4.8460	4.5640	4.0260
Motivación y entusiasmo del profesor	4.6670	4.6410	4.0770
Motivación que despierta en los alumnos	4.2560	4.1030	3.6920
Organización y estructuración de la exposición	4.3590	3.9490	3.2560
Interacción profesor-alumno en clase	4.1540	3.9230	4.0000
Relación afectiva profesor-alumno.	3.4620	3.7180	3.7690
VALORACION GLOBAL	4.2910	4.1500	3.8030

Esta nueva valoración cambia los resultados hasta un máximo de un 7%. En este caso, los profesores 1 y 3 mejoran en 4 de las dimensiones y en la valoración global. El profesor 2 empeora en 4 dimensiones y en la valoración global.

En la Tabla N° 4 figuran los resultados de los tres profesores en la valoración inicial y en los dos criterios de calidad evaluativa a través de los cuales se modificó su puntuación inicial.

TABLA Nº 4

	PROFESOR 1			PROFESOR 2			PROFESOR 3		
	TODOS	50%	66.7%	TODOS	50%	66.7%	TODOS	50%	66.7%
Conocimientos y competencia	4.8125	4.8360 +0.49%	4.8460 +0.70%	4.5625	4.5520 -0.23%	4.564005 +0.03%	3.9750	3.9550 -0.50%	4.0260 +1.28%
Motivación y entusiasmo del profesor	4.5750	4.5970 +0.48	4.6670 +2.01%	4.6625	4.6720 +0.20%	4.641000 -0.46%	4.0500	4.0300 -0.49%	4.0770 +0.67%
Motivación que despierta en los alumnos	4.2625	4.2540 -0.20%	4.2560 -0.15%	4.2250	4.1790 -1.09%	4.103005 -0.89%	4.6250	3.6120 -0.36%	3.6920 +1.85%
Organización y estructuración de la exposición	4.1750	4.2540 +1.89%	4.3590 +4.41%	3.9375	3.9250 -0.32%	3.949002 +0.29%	3.2625	3.2690 +0.20%	3.2560 -0.20%
Interacción profesor-alumno en clase	4.2000	4.1940 -0.14%	4.1540 -1.10%	4.1875	4.1490 0.92%	3.923000 6.32%	4.0000	3.9850 -0.38%	4.0000 =
Relación afectiva profesor-alumno	3.4250	3.5220 +2.83%	3.4620 +1.08%	4.0125	3.9850 -0.69%	3.78005 7.34%	3.7750	3.7460 -0.77%	3.7690 -0.16%
VALORACION GLOBAL	4.2417	4.2760 +0.81%	4.2910 +1.16%	9.2646	4.2440 -0.48%	4.150001 -2.69%	3.7813	3.7660 -0.40%	3.8030 +0.57%

El análisis por profesor demuestra que el profesor 1, a medida que aumenta la calidad evaluativa del alumnado, mejora en cuatro de las dimensiones y en la valoración global. En general, podemos afirmar que fue inicialmente “infravalorado”.

Al profesor 2 prácticamente le sucede lo contrario, a medida que mejora la calidad evaluativa del alumnado, empeora en 4 de las dimensiones evaluadas y en la valoración global. En general, podemos afirmar que fue “sobrevalorado”.

El profesor 3 muestra un perfil diferente. En un primer momento, empeoran sus puntuaciones ligeramente y luego en la máxima calidad evaluativa mejora su valoración inicial, pero sin grandes diferencias. Podríamos afirmar que fue “justamente” evaluado desde el principio.

Discusión y Conclusiones

En este tipo de evaluaciones no existe un referente de evaluación objetiva, puesto que no tenemos una medida objetiva de la valoración que realmente merece cada profesor. La media tampoco es un referente objetivo ajeno a los alumnos. La media depende de todos los alumnos, buenos o malos evaluadores. Por lo tanto, también está condicionada. Sin embargo, el referente de no alejarse permanentemente de lo que piensan la mayoría del grupo es el único referente de calidad evaluativa, que con cierta objetividad podemos manejar. Nótese que no exigimos estar siempre de acuerdo con la media del grupo, sino «no estar siempre en desacuerdo con la valoración media del grupo. Y que este estar de acuerdo o en desacuerdo está representado por un intervalo amplio, más/menos una desviación típica, es decir, estar de acuerdo con el 68% de los compañeros/as.

En definitiva, esta fórmula para depurar los resultados es factible y parece lógico eliminar, como se suele hacer en esos procesos evaluativos (jueces de algunos deportes), aquellas puntuaciones extremas. Para ello bastaría con:

- 1) Transformar las valoraciones de los alumnos en puntuaciones típicas en función de la media y la desviación típica del grupo.
- 2) Señalar, vistas las puntuaciones Z de toda la población evaluadora, un criterio de corte que indique qué puntuaciones serán extremas y cuáles no, es decir, fijar un porcentaje de alumnos a eliminar, o dicho de otra manera, un criterio de calidad evaluativa.
- 3) Calcular la valoración media de cada profesor con la totalidad del alumnado y luego sólo con los alumnos con puntuaciones no extremas (no eliminados).
- 4) Eliminar las valoraciones de los alumnos extremos antes de dar los resultados por profesor.

Bibliografía

- ABRAMI, P., LEVENTHAL, L., PERRY, R. y BREEN, L. (1976). Course evaluation. How? *Journal of Educational Psychology*, 68, 300-304.
- ABRAMI, P. C. (1989). How should we use student ratings to evaluate teaching. *Research in Higher Education*, 30, 2; pp. 221-227.
- BAEZ, B. F., GUARRO, A., MIRO, M. T., PEÑATE, W., PEREZ, F. y SAN LUIS, C. (1996). *Evaluación y Mejora de la Calidad Docente en la Universidad de La Laguna. Una propuesta de actuación*. Comisión Técnica de Evaluación de la Calidad Docente. La Laguna: Universidad de La Laguna.
- BAXTER, E. P. (1991). The TEVAL Experience, 1983-1988. The impact of a student evaluation of teaching scheme on university teachers. *Studies in Higher Education*, 16; 151-178.
- CAHN, S. M. (1987). Faculty Members Should Be Evaluated by Their Peers, Not by Their Students. *The Chronicle of Higher Education*, nº 14.
- CARRIER, N., HOWARD, G., y MILLER, W. (1974). Course evaluation. When? *Journal Educational Psychology*, 66, 609-613.
- CASTRO, J. J., HERNÁNDEZ, P. y GARCIA, L. A. (1996). *Factores moduladores de la evaluación del profesor*. Tesis doctoral, Universidad de La Laguna.
- COHEN, P. A. (1981). Student ratings of instruction and student achievement. A meta-analysis of multisection validity studies. *Review of Educational in Education*, 3, 4; pp. 281-309.
- COOK, S. S. (1989). Improving the quality of student ratings of instruction. A look at two strategies. *Research in Higher Education*, 30, 1; pp. 31-45.
- De MIGUEL, M. y RODRÍGUEZ, S. (1990). Evaluating university system. a Spanish faculty point of view. Ponencia presentada en el *12th Annual Forum of the European Association for Institutional Research*, Lyon, Septiembre.
- DIAZ-AGUADO, M. J. (1986). Interacción Profesor-Alumno. En P. HERNANDEZ (Dir.). *Psicología de la Educación y Enseñanza del Universitaria*. Instituto de Ciencias de la Educación. La Laguna: Universidad de La Laguna.
- (1986). Mesa Redonda. Experiencias sobre Evaluación del Profesorado Universitario. En P. Hernández (Dir.). *Psicología de la Educación y Enseñanza del Universitaria*. La Laguna: Instituto de Ciencias de la Educación de la Universidad de La Laguna.
- (1987). La percepción del profesor por el alumno. Expectativas y actitudes. En J. BELTRAN (Ed.). *Psicología Educativa*. Volumen I. Madrid: UNED.
- (1987). Eficacia del profesor. En J. BELTRAN (Ed.). *Psicología Educativa*. Volumen I. Madrid: UNED.
- DOWELL, D. y NEAL, J. (1982). A Selective Review of the Validity of Student Rating of Teaching. En *Journal of Higher Education*, 53; 51-62.
- ENTWISTLE, N. y TAIT, H. (1990). Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments. *Higher Education*, 19; pp. 169-194.
- ENTWISTLE, N. J. y RAMSDEN, P. (1983). *Understanding Student Learning*. London: Croom Helm.
- ESCUADERO, T. (1993). Enfoques Modélicos en la evaluación de la enseñanza Universitaria. En *III Jornadas Nacionales de Didáctica Universitaria. Evaluación y Desarrollo Profesional*. Las Palmas: Servicio de Publicaciones de la ULPGC.
- 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, 2; pp. 139-213.
- (1989a). Instructional effectiveness of college teachers as judge by teachers themselves, current and former students, colleagues, administrators and external (neutral) observers. *Research in Higher Education*, 30, 2; pp. 137-194.
- (1989b). The association between student ratings of specific instruncional dimensions and student achievement. refining and extendling the synthesis of data from multisection validity studies. *Research in Higher Education*, 30, 6; pp. 583-645.
- FIRTH, M. (1979). Impact of work experience on the validity of student evaluation of teaching effectiveness. *Journal of Educational Psychology*, 71, 5726-30.
- FRANKE-WIKBERG, S. (1990). Evaluating Education Quality on the Institutional level. *Higher Education Management*, 2; 271-292.

- GOOD, T. L. y MULRYAN, C. (1990). Teacher ratings. A call for teacher control and self-evaluation. En J. Millman Y L. Darling-Hammond (Ed.). *The New Handbook of Teacher Education*. Newbury Park (Cal.): Sage Publications; pp. 191-215.
- HERNÁNDEZ, P. (1986). Mesa Redonda. Experiencias sobre Evaluación del Profesorado Universitario. En P. Hernandez (Dir.). *Psicología de la Educación y Enseñanza del Universitario*. La Laguna: Instituto de Ciencias de la Educación de la Universidad de La Laguna.
- HOFFMAN, J. y KREMER, L. (1980). Attitudes toward higher education and course evaluation. *Journal of Educational Psychology*, 72, 610-617.
- JORNET, J., VILLANUEVA, P., SUAREZ, J. y ALFARO, I. (1989). Proyecto de implantación de un sistema de evaluación del profesorado en la Universidad de Valencia. En *Informes de investigación evaluativa, nº 1*, Consideraciones metodológicas sobre evaluación y mejora de la docencia universitaria. Valencia: Universidad de Valencia.
- KELLEY, A. (1972). Uses and abuses of course evaluation as measures of educational output. *Jour. Econ. Educ.*, 4, 13-18.
- KURTZ, R., MUELLER, J. J., GIBBONS, J. L. y DICATALDO, F. (1989). Faculty Suggestions for the refinement of the Concept and its Measurement. *Journal of Higher Education*, 60; 43-58.
- LODER, C. P. J. (Ed.) (1990). *Quality assurance and accountability in higher education*. Londres: Kogan Page.
- MARSH, H. W. (1987). Student evaluation of teaching. En M. J. Dunkin (Ed.). *The International Encyclopedia of Teaching and Teacher Education*. Oxford: Pergamon Press. pp. 181-187.
- MARSH, H. W. y ROCHE, L. A. (1992). The use of student evaluation of university teaching in different setting, the applicability paradigm. *Australian Journal of Education*, 36, pp 278-300.
- MARSH, H. (1982). Validity of students evaluations of college teaching. a multitrait-multimethod analysis. *Journal of Educational Psychology*, 74, 265-279.
- MARSH, H., OVERAL, J. Y KESLER, P. (1979). Validity of students evaluations of instructional effectiveness. A comparison of faculty self evaluation and evaluation by their students. *Journal of Educational Psychology*, 71, 149-160.
- MARSH, H. W., TOURON, J. y WHEELER, B. (1985). Students' evaluations of university instructors. The applicability of american instruments in a spanish setting. *Teaching & Teacher Education*, 1, 2; pp. 123-138.
- MENGES, R. J. y BRINKO, K. T. (1986). Effects os Student Evaluation Feedback. A meta-Analysis of Higher Education Research. Ponencia en el *Annual Meeting of the A.E.R.A.*, San Francisco.
- MURRAY, H. G. (1984). The impact of formative and summative evaluation of teaching in North American Universities. *Assessment and Evaluation in Higher Education*, 9, 12; 117-132.
- OVERAL, J y MARSH, H. W. (1980). Students' evaluations of instruction. a longitudinal study of their stability. *Journal of Educational Psychology*, 72, 3, 321-325.
- RAMSDEN, P. (1991). A Performance Indicator of Teaching Quality in Higher Education. The experience Questionnaire. *Studies in Higher Education*, 16; 129-150.
- RANSDEN, P. y ENTWISTLE, N. J. (1981). Effects of Academic Department on Students Approaches to Studying. *British Journal of Education Psychology*, 51; 368-383.
- ROGERS, C. (1987). *Psicología Social de la Enseñanza*. Madrid: Centro de publicaciones del M.E.C. y Visor Libros.
- RODRIGUEZ, S. (1989). La evaluación del profesorado universitario. Análisis de la opinión del profesorado de la Universidad de Barcelona. En *Informes de Investigación Evaluativa nº 1*. Consideraciones metodológicas sobre la evaluación y mejora de la docencia universitaria. Universidad de Valencia; pp. 211-235.
- RODRIGUEZ, S. (1991). Dimensiones de la calidad universitaria. Ponencia presentada en el *I Congreso Internacional sobre calidad de la educación universitaria*, 6-8 de marzo. Puerto de Santa María (Cádiz).
- STONE, E., RABINOWITZ, S. y SPOOL, M. (1977). Effect of anonymity on student evaluations of faculty performance. *Journal of Educational Psychology*, 69, 274-280.
- TOURON, J. (1989). La validación de constructo. su aplicación al CEED (Cuestionario para la mejora de la eficacia docente). *Revista Bordón*, 41, 4; pp. 735-756.
- VILLAR, L. M. (1987). Evaluación dela conducta docente y del clima psicosocial de la clase. En *Consideraciones metodológicas sobre Evaluación y Mejora de la Docencia Universitaria*. Servei de Formació Permanent de la Universitat de Valencia.
- WINGINGTON, H., TOLLEFSON, N. y RODRIGUEZ, E. (1989). Students ratings of instructors revisited. Interactions among class and instructor variables. *Research in Higher Education*, 30, 3; pp. 331-344.