

Valoración de la calidad docente:

El profesorado.

Un modelo de evaluación circular

Juan Fernández

UCM

EDITORIAL
COMPLUTENSE

Queda rigurosamente prohibida sin la autorización escrita de los titulares del Copyright, bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamo público.

© 2008 by Juan Fernández Sánchez
© 2008 by Editorial Complutense, S. A.
Donoso Cortés, 63 – 4. planta (28015) Madrid
Tels.: 91 394 64 60/1 Fax: 91 394 64 58
e-mail: ecsa@rect.ucm.es
www.editorialcomplutense.com

Primera edición: noviembre 2008

ISBN: 978-84-7491-943-1

VALORACIÓN DE LA CALIDAD DOCENTE: EL PROFESORADO UN MODELO DE EVALUACIÓN CIRCULAR

JUAN FERNÁNDEZ

Juan Fernández

Catedrático de Psicología Evolutiva y de la Educación en la Facultad de Psicología de la Universidad Complutense de Madrid. Inicia su **programa de investigación** sobre la *evaluación de la calidad docente* en los inicios de la década de los 80 y lo continúa en nuestros días. *Ha realizado evaluaciones para distintas instituciones públicas y privadas* (universidades, institutos, ayuntamientos, sociedades médicas,...), que han dado lugar a **cientos de informes técnicos**. Ha publicado sobre estos asuntos en diversas revistas estadounidenses (*Educational and Psychological Measurement, Sex Roles...*) y españolas (*Psicothema...*). Su modelo de evaluación ha visto la luz en la revista de la Organización para la Cooperación y el Desarrollo Económico (OCDE): *Higher Educational Management*. Algunos de sus instrumentos de evaluación han sido traducidos a *distintos idiomas* y utilizados en diversas *universidades de varios países*. Ha impartido conferencias sobre esta materia en España, en países europeos e iberoamericanos.

- Fernández, J. (1997). A Spanish evaluation model in higher education: Circular evaluation. *Higher Education Management, 9*, 71-84.
- Fernández, J., & Mateo, M. A. (1992). Student evaluation of university teaching quality: Analysis of a questionnaire for a sample of university students in Spain. *Educational and Psychological Measurement, 52*, 675-686.
- Fernández, J., & Mateo, M. A. (1993). The development and factorial validation of the Academic Setting Evaluation Questionnaire. *Educational and Psychological Measurement, 53*, 425-435.
- Fernández, J., & Mateo, M. A. (1994). The circular evaluative process of higher education: A Spanish sample. In OECD (Ed.), *Evaluation and the decision making process in higher education: French, German, and Spanish experiences* (pp. 83-94). Paris: OECD Publications.
- Fernández, J., & Mateo, M. A. (1998). Student and faculty gender in ratings of university teaching. *Sex Roles, 37*, 997-1003.
- Fernández, J., Mateo, M. A. y Muñiz, J. (1996). Valoración por parte del profesorado de la evaluación docente realizada por los alumnos. *Psicothema, 8*, 167-172.
- Fernández, J., Mateo, M. A., & Muñiz, J. (1995). Evaluation of the academic setting in Spain. *European Journal of Psychological Assessment, 11*, 133-137.
- Mateo, M. A. y Fernández, J. (1992). Análisis confirmatorio de la estructura dimensional de un cuestionario para la evaluación de la calidad de la enseñanza. *Investigaciones Psicológicas, 11*, 73-82.
- Mateo, M. A. y Fernández, J. (1993). Dimensiones de la calidad de la enseñanza universitaria. *Psicothema, 5*, 265-275.
- Mateo, M. A., & Fernández, J. (1995). Evaluation of the setting in which university faculty carry out their teaching and research functions: The ASEQ. *Educational and Psychological Measurement, 55*, 329-334.

Correo electrónico: jfernandez@psi.ucm.es

Tfno.: 91 3943162

Dirección: Departamento de Psicología Evolutiva y de la Educación

Facultad de Psicología

Campus de Somosaguas

28223 – Madrid (España)

VALORACIÓN DE LA CALIDAD DOCENTE: EL PROFESORADO UN MODELO DE EVALUACIÓN CIRCULAR

JUAN FERNÁNDEZ

Catedrático de Psicología Evolutiva y de la Educación

Facultad de Psicología. Universidad Complutense de Madrid

Índice

	Pág.
Prólogo	8
PRIMERA PARTE: CONCEPTUAL	
Capítulo primero	
La cultura de la evaluación dentro del ámbito educativo	
1. El desglose del ámbito educativo	13
1.1. Profesores, alumnos y currículos	13
1.2. Evaluaciones, asesoramientos, intervenciones y contextos	17
2. Los componentes básicos de la denominada cultura de la evaluación	18
2.1. Etapas de la evaluación	19
2.2. Institucionalización de la evaluación	21
2.3. Funciones del evaluador	23
2.4. Plaga o remedio para la mejora de lo evaluado	24
3. Comentarios	26
Capítulo segundo	
Evaluabilidad, evaluación y metaevaluación	
1. Valoraciones de la evaluabilidad	29
2. Las evaluaciones	31
2.1. Tipos de evaluaciones	33
2.2. Principales enfoques	35
3. Las metaevaluaciones	41
4. Comentarios	44

Capítulo tercero

Las evaluaciones centradas en el profesor a lo largo de su carrera profesional

1. Introducción	46
2. Los fines de la evaluación	47
3. Aspectos a evaluar	47
4. Evaluación del profesorado en sus distintas etapas profesionales	48
5. Distintos procedimientos de evaluación	52
5.1. La observación	53
5.2. La autoevaluación	54
5.3. El portafolios/archivo o carpeta docente del profesor	55
5.4. Las opiniones de los alumnos	56
5.5. Otros procedimientos	59
6. Comentarios	60

SEGUNDA PARTE: OPERATIVA

Capítulo cuarto

El modelo de evaluación circular

1. Introducción	62
2. Las opiniones de los alumnos	65
2.1. Las variables relevantes frente a las irrelevantes	65
2.2. Elección o elaboración de instrumentos	67
2.3. La devolución de la información	69
2.4. Los criterios de evaluación: el de la escala de valoración, el heterocomparativo y el autocomparativo	71
3. Las opiniones de los profesores	73
3.1. La elaboración de un instrumento específico de evaluación	73

3.2. La circularidad	84
4. La metaevaluación	85
5. Algunas cuestiones en torno a posibles sesgos	88

TERCERA PARTE: LO CONCEPTUAL Y LO OPERATIVO

Capítulo quinto

Síntesis de lo conceptual/operativo

1. La cultura de la evaluación y la utilización de los resultados de la evaluación	94
2. El proceso evaluador	95
3. Los diversos procedimientos de evaluación	97
4. Los enfoques integradores	99
5. La evaluación, compañera de viaje de cualquier profesor	100
6. La evaluación del profesor en el contexto de las demás evaluaciones	101
Epílogo metaevaluativo: el libro que no quise escribir	105
Referencias	107
Anexos:	
1. Dimensiones del Profesorado Universitario: DPU	116
2. El instrumento de valoración de los alumnos: CEPUC-29(CUTEQ)	118
3. El instrumento de valoración de los profesores: CECA-59(ASEQ)	121
4. Devolución de la información en el modelo de evaluación circular	126
4.1. La evaluación de los alumnos	127
4.2. La evaluación de los profesores	138
5. El instrumento de autoevaluación del profesorado: AUPRO	143
Índice de autores	146

Prólogo

Esta obra se enmarca dentro de lo que hoy denominamos como *sociedad del conocimiento y de la información* y, por tanto, del aprendizaje, sin olvidarnos de la otra cara de la misma moneda, enseñanza/educación, que necesariamente han de prolongarse a lo largo de todo el ciclo vital. Este hecho ya nos está marcando un gran cambio para los individuos, instituciones educativas y sociedades del siglo XXI: el desarrollo, el aprendizaje y la enseñanza son fenómenos consustanciales al ser humano que lo marcarán y condicionarán, posiblemente más que ningún otro, en todas y cada una de las etapas de su vida. El aprendizaje y la enseñanza no son, por consiguiente, factores típicos de un determinado periodo de la vida (la infancia y la adolescencia), como a lo largo de buena parte del siglo XX se pensó, debido ante todo y sobre todo a la influencia de ciertos autores y manuales especializados en la psicología del desarrollo.

Los procesos de enseñanza/aprendizaje, dentro de una sociedad del conocimiento, van a incorporar una serie de nuevos elementos –todos los asentados y derivados de Internet, así como de las nuevas tecnologías (los variados soportes lógicos)- que obligarán a profesores, alumnos, padres, autoridades académicas y sociedad en cuanto tal a cambiar buena parte de sus concepciones clásicas y actuales sobre el aprendizaje y la enseñanza. La literaria Biblioteca de Babel borgesiana va ir abandonando cada vez más el terreno de la imaginación para entrar en el campo de nuestra vida cotidiana, por lo que la realidad sociocultural sobrepasará, una vez más, al producto imaginario de una de las mentes más cultivadas del pasado siglo. Ante este hecho insólito se requiere un *mediador* –el profesor- que haga consciente al discente – en sus diversos contextos: formal, informal y no-formal- de que la acumulación de conocimientos ya no puede ser el objetivo último de su docencia (pues se halla presente y accesible en la biblioteca de Babel informatizada, de forma instantánea, desde cualquier parte del mundo), cuanto el *conocimiento significativo* y, por ende, funcional. Si la meta es el conocimiento, que no mera acumulación de conocimientos, el profesor ha de ayudar al alumno a *seleccionar y discriminar* (función capital del buen docente a lo largo de todos los tiempos, pero muy especialmente en la actualidad y sobre todo en el futuro) a aquellos autores y obras que mejor puedan desarrollar sus *capacidades*, sus *competencias*, sus *valores* y sus *sistemas de creencias*. Ahora bien, esta mediación ha de ser muy distinta en función de cada etapa del aprendiz: muy estructurada en la infancia, menos en la adolescencia y todavía menos en la edad adulta, para volver a una mayor estructuración en la vejez, aunque matizada en función de las características diferenciales de los estudiantes: más estructurada para los alumnos *analíticos* y *prácticos*, en el decir sternbergiano, así como para aquellos sujetos con ciertas necesidades educativas especiales, y menos estructurada, lógicamente, para los más *creativos* e independientes (Sternberg y Spear-Swerling, 1996).

El objetivo básico de este libro es ofrecer un material especialmente seleccionado para que podamos saber hasta qué punto el profesorado, en los distintos niveles educativos, cumple adecuadamente con esta función de *mediador competente, de auténtico maestro, de modelo digno de ser imitado*. Este material se ha dividido en tres grandes conglomerados que vertebran todo el libro: uno dedicado a la clarificación conceptual; otro de tipo operativo, con el fin de que todo lo conceptual se vea adecuadamente materializado; y, por fin, uno de conclusiones, con la pretensión de que éstas puedan cobrar vida en el terreno de la realidad cotidiana de los diferentes contextos educativos, si así lo estima oportuno y lo desea el lector.

Cada uno de estos conglomerados se desglosa, a su vez, en los correspondientes capítulos. En el primero, se trata de ofrecer una síntesis, lo más coherente posible, de la *evaluación* dentro de un contexto muy concreto: *el educativo*. Se especifican en este capítulo los elementos esenciales del ámbito educativo: profesores, alumnos, currículos y demás agentes, partiendo de la base de que todos ellos pueden y deben ser objetos y sujetos de evaluación. Esta evaluación debiera estar encaminada principalmente a la consecución de la excelencia académica, entendiendo por tal el máximo desarrollo intelectual, personal y social de los agentes que participan en un contexto determinado de aprendizaje y enseñanza, asumiendo en todo momento que el marco social actual es el de la sociedad del conocimiento. Todos los miembros de la comunidad de aprendizaje y enseñanza van a necesitar la evaluación si su objetivo prioritario es el desarrollo profesional y la calidad del sistema de aprendizaje/enseñanza. De ahí la necesidad de una cultura de la evaluación, como paso previo obligado de cualquier programa cuyo fin sea el mejoramiento del sistema educativo.

Ahora bien, es necesario ir más allá de dicha síntesis con el fin de poder captar los componentes conceptuales esenciales del *proceso evaluador* (capítulo segundo). Se analizarán aquí, por tanto, cada uno de estos conceptos y sus implicaciones, para poder integrar los elementos dispersos que aparecen en diferentes enfoques de evaluación. Gracias al análisis de la *evaluabilidad*, nos daremos cuenta de que es preciso ser mucho más concretos a la hora de ponernos metas u objetivos educativos. Los máximos responsables políticos y académicos suelen utilizar frecuentemente conceptos cargados de connotaciones positivas (excelencia docente; calidad de la enseñanza; desarrollo profesional, de la inteligencia, creativo, “humanístico”; aprendizaje cooperativo; enseñanza de valores, etcétera), pero la dura realidad nos obliga a preguntarnos sobre cómo es posible concretar y materializar todo ello dentro de cada contexto de aprendizaje/enseñanza: educación infantil, primaria, secundaria, universidad y educación permanente. Ése sería el núcleo duro de la evaluabilidad. No es, pues, por azar por lo que este concepto ha sido tan poco estudiado y mucho menos puesto en práctica. De hecho, debido a su ausencia o invisibilidad, una buena parte de la responsabilidad política y académica queda diluida en la ambigüedad de unos objetivos casi místicos que no pueden ser evaluados y, por tanto, imposibilitan, o al menos dificultan, acciones bien fundamentadas sobre las decisiones tomadas por estas autoridades. No es de extrañar, por otra parte, que este tipo de objetivos sea precisamente el denominador común de la mayoría de las ideologías, no importa su concreción política, dadas las connotaciones positivas de los mismos y de su gran valor para la captación fácil del voto.

Tras la evaluabilidad se abre paso la evaluación, o mejor dicho, los *distintos tipos de evaluación*: inicial o diagnóstica, de procesos y de productos. Además, parece conveniente que, tras tener en mente las decenas de definiciones o concepciones de evaluación, tratemos de alcanzar un cierto denominador común de todas ellas, incluso cuando se aplican a objetos tan distintos como los alumnos, los profesores, el currículum, o a programas, proyectos o contextos educativos.

Tras la evaluabilidad y la evaluación aparece la *metaevaluación*. Este concepto ha sido también mucho menos estudiado que el de la evaluación y, sin embargo, está llamado a cumplir funciones relevantes en cualquier proceso evaluador. De hecho a él hemos de recurrir para dilucidar rigurosamente tanto el impacto –positivo o negativo,

deseado o indeseado, previsto o imprevisto- como el alcance de las evaluaciones. A veces, ingenuamente, se parte del supuesto de que las evaluaciones siempre han de aportar algo positivo, pero sólo gracias a la metaevaluación es posible asegurarse de que así ha sido.

Teniendo todo esto en cuenta, es posible adentrarse ya en lo que es el núcleo fundamental de nuestro interés: la *evaluación del profesorado*, poniendo desde un comienzo de manifiesto que la finalidad principal de la evaluación va a ser el *desarrollo profesional* (capítulo tercero). Se da cuenta en este capítulo de los distintos procedimientos de evaluación (observación, grabación, portafolios, autoevaluación, evaluación por pares, por expertos, por autoridades, etcétera), destacando el más utilizado hasta la fecha en los contextos universitarios: la evaluación a través de las opiniones de los alumnos. Es sin duda el más investigado y el mejor contrastado, por lo que es posible analizar detalladamente sus ventajas y desventajas. Gracias a este abundante *corpus* bibliográfico, este procedimiento nos puede servir, además, de modelo para ver las pruebas a que han de ser sometidos los demás procedimientos hasta alcanzar una credibilidad científica suficiente como para que puedan ser utilizados con cierta garantía de éxito.

Aunque el objetivo prioritario de la evaluación, no importa cuál pueda ser el procedimiento, será el desarrollo profesional, se tendrán en cuenta también otros tipos de objetivos, muy especialmente el referido a la selección, promoción y cuantía salarial del profesorado estable.

A la hora de poder llevar a cabo ese proceso evaluador se cuenta hoy en día con numerosos enfoques, cada uno de ellos con sus puntos fuertes y débiles. En el libro se ofrecerá una explicación bastante exhaustiva de uno de ellos: el denominado *modelo de evaluación circular*. Este modelo ha sido desarrollado por el autor a lo largo de las dos últimas décadas del siglo XX, gozando en estos momentos de una bastante buena acogida nacional e internacional (capítulo cuarto). Se especificarán todos y cada uno de los componentes de este modelo: evaluación del alumnado, del profesorado, de la relación entre ambos, del currículum, del contexto de aprendizaje/enseñanza, de las autoridades académicas y políticas y, por fin, se tendrá en cuenta la metaevaluación, como una de las partes integrales del proceso evaluador. Se mostrará, tanto teórica como prácticamente, la capacidad que tiene el modelo para integrar aspectos que aparentemente resultan contrapuestos desde otros enfoques. Tal vez su valor más destacado resida en el *cruce de la información proveniente de los distintos agentes que participan en cada contexto escolar*. Cada uno de ellos es a la vez sujeto y objeto de evaluación y cada uno de ellos nos puede y debe ofrecer su visión (aquella para la que está más capacitado) sobre lo que está sucediendo en un determinado ámbito educativo.

Finalmente, se presentarán unos comentarios y conclusiones, que entiendo pueden ser útiles para una buena evaluación del profesorado en cualquiera de los niveles educativos (capítulo quinto). Lo que se pretende en este último apartado es realizar una síntesis, mínimamente coherente, de todos aquellos aspectos que hayan resultado más relevantes en cada uno de los capítulos, pertenecientes tanto al conglomerado de lo conceptual como al de lo operativo, junto con sus implicaciones. Los dos hilos conductores para esta síntesis han sido, por un lado, la perspectiva teórica (se pretende que el lector, gracias a la asimilación de lo referido en el libro, cuente con una visión del proceso evaluador garantizada científicamente) y, por otro, la perspectiva práctica, de

forma que, gracias a lo leído, se sienta capacitado para emprender algún proceso de evaluación, bien referido a sí mismo, o bien al profesorado con el que se relaciona.

Como se podrá comprobar, este libro es fruto tanto de las lecturas seleccionadas del autor como, sobre todo, de su propia línea o programa de investigación, que se inicia a comienzos de los 80 y que dura hasta nuestros días. De ahí las continuas referencias, sobre todo a lo largo del segundo conglomerado (capítulo cuarto), a trabajos propios ya publicados; trabajos que unas veces son de tipo teórico pero que fundamentalmente dan cuenta de investigaciones empíricas realizadas en España durante los últimos 25 años. Esto ha permitido aligerar considerablemente esta obra, pues el lector siempre podrá recurrir a esos trabajos citados, bien para ampliar más la información o bien para verificar los aspectos más técnicos de la misma: diversos tipos de análisis de datos, propiedades formales de los distintos instrumentos de evaluación, entre otros.

He intentado en todo momento que el libro pueda ser leído sin dificultad, que goce de solidez científica y que, tras su lectura, se pueda pasar a la ejecución de la práctica profesional de la evaluación, en este caso, de la evaluación del profesorado. He utilizado, en ocasiones, la primera persona del singular y, en otras, la primera del plural. Esto es debido a que, a veces, hago referencia a lo realizado básicamente por mí, mientras que en otros lugares comento trabajos llevados a cabo por el equipo de investigación que vengo coordinando desde hace tiempo. El lector, ahora evaluador, tiene la última palabra.

Pero, ¿de qué lector estoy hablando? ¿En quién he estado concretamente pensando al escribir el libro? En las autoridades académicas preocupadas por los asuntos de la evaluación del profesorado, en los profesores de los distintos niveles educativos (infantil, primaria, secundaria y universidad) y en los que imparten su docencia en diversos contextos de educación no formal o informal. Por supuesto, en los profesionales de la evaluación, así como en los gestores de procesos y productos educativos. Y, en general, en todos aquellos que desearían contar con algunas nociones sólidas sobre este amplio campo, tan de moda en nuestros días, como es el de la evaluación de la calidad.

Dos son las personas que generosamente han dedicado su tiempo a una lectura pormenorizada del manuscrito antes de que éste pudiese ver la luz pública. Su objetivo, a petición mía, fue el de detectar los posibles errores o incoherencias conceptuales, así como las posibles faltas formales. Mucho ha ganado el libro, en ambos aspectos, tras sus comentarios. No obstante, la última responsabilidad ante cualquier falta de uno u otro tipo, como es obvio, es únicamente mía. Mi agradecimiento, pues, sincero y de corazón a Paloma Rodríguez y a Miguel Ángel Mateo por su inteligente esfuerzo crítico y por su desbordante generosidad.

CAPÍTULO PRIMERO

La cultura de la evaluación dentro del ámbito educativo

1. El desglose del ámbito educativo

Parece conveniente que si se va a hablar de la evaluación del profesorado, trate primero de dejar claro lo que a partir de ahora entenderé como la especificidad del ámbito educativo para cualquiera de los distintos niveles en que suele dividirse la educación formal internacional: educación infantil, primaria, secundaria, universidad y educación permanente, allá donde ésta tenga ya o pueda tener lugar en el futuro. Los principales protagonistas de este ámbito han sido en todas las sociedades, por un lado, los alumnos, por el otro, los profesores y, como cordón umbilical que une necesariamente a ambos agentes, los currículos (véase Figura 1).

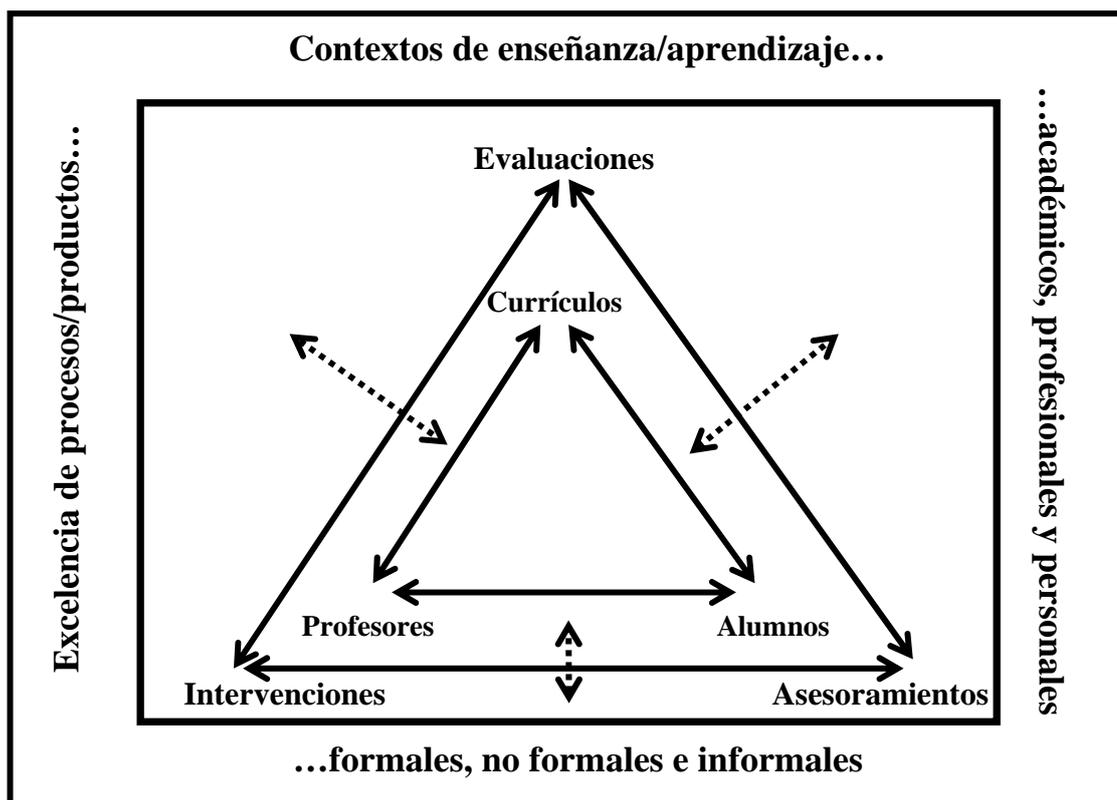


Figura 1. Especificidad del ámbito educativo (Fernández, 2001).

1.1. Profesores, alumnos y currículos

Hasta aproximadamente finales del siglo XIX la principal función del profesor era la transmisión de conocimientos. Éstos no estaban al alcance más que de una minoría de la población y esa minoría eran precisamente los profesores. A éstos, las diferentes sociedades les encargaban la importante labor de hacer de correa de transmisión de la sabiduría (religiosa, filosófica, científica, técnica y artística) y de la cultura (valores) que la especie humana había ido acumulando a lo largo de los siglos. Hasta esas fechas el presupuesto subyacente a la misión encomendada al profesorado era muy simple: cuantos más conocimientos poseyera el profesor mejor para su sociedad, pues de esta forma más contenidos podría transmitir a sus alumnos. La peor deficiencia de un profesor habría de ser, en consecuencia, la escasez o limitación en el correspondiente *corpus* de conocimientos, puesto que tal limitación necesariamente se reflejaría en sus alumnos. Desde este modelo, que denominaré a partir de ahora clásico,

la principal función de todo profesor estaba claramente determinada desde un principio –la *transmisión de conocimientos*–, otorgándole la sociedad un poder y autoridad suficientes, acompañados ambos incluso de la opción a la coerción física, para que los pudiese ejercer casi con total libertad sobre sus alumnos. El papel del profesor era lógicamente activo en todo momento: él era el que establecía el cuánto de lo que había que aprender, el cuándo, el cómo y demás condicionantes del aprendizaje, que por la escasez e inaccesibilidad de medios tenía que ser básicamente memorístico. Este papel activo del profesor se correspondía a la perfección con el papel esencialmente pasivo del alumno en todo lo tocante al diseño curricular: su tarea era callar y aprender de memoria lo que el profesor le exigía en cada momento.

Es este modelo educativo simplista (me refiero a su presupuesto básico: el de la transmisión de conocimientos desde el profesor a los alumnos) el que va a heredar uno de los enfoques psicológicos que más influencia ha tenido a lo largo de la primera parte del siglo XX: *el conductismo*. Su labor básica ha sido la de ofrecer una tecnología eficaz a la hora de realizar esa transmisión de conocimientos entre profesores y alumnos. Para ello se contó con la experiencia previa de los logros espectaculares obtenidos con diversos tipos de animales. Gracias a estos experimentos fue posible inferir las así llamadas “leyes” o principios de aprendizaje. Si eran eficaces con los animales con más razón habrían de serlo con los humanos. De hecho, así lo manifiesta contundentemente uno de los fundadores de esta corriente. Watson (1926), reta a que se le proporcione una docena de niños sanos de entre los que elegiría uno al azar. Gracias únicamente a la pertinente educación por él proporcionada, sin importar ni las aptitudes ni las inclinaciones del sujeto ni la raza de sus ancestros, haría de él un auténtico especialista en cualquier materia: médico, abogado, artista, empresario o incluso un mendigo o un ladrón. Obviamente, tal reto estaba fundamentado en su concepción del niño como *tabula rasa*, en el modelo del condicionamiento clásico y en su visión del educador como mero transmisor de conocimientos, todo ello funcionando dentro de una contingencia que él podía controlar plenamente mediante los oportunos refuerzos y castigos (positivos o negativos, en ambos casos), en función del mayor o menor acercamiento conductual del sujeto a la meta establecida.

Así cada profesor se debía convertir, científicamente hablando, en el administrador de refuerzos y castigos, según una contingencia preestablecida, en tanto armas que se habían mostrado poderosas a la hora de conseguir el éxito de los aprendizajes en animales y personas. Es indudable que la aplicación de las teorías y técnicas conductistas (condicionamiento clásico y condicionamiento operante, básicamente), en el ámbito educativo, han supuesto un avance en nuestros conocimientos actuales sobre todo lo relacionado con el aprendizaje en las aulas. De ahí que se cuente hoy en día con una bibliografía abundante a disposición de especialistas y del propio profesorado (véanse Bergan, 1982 o Schunk, 1991, como ejemplos ilustrativos de lo afirmado).

Ahora bien, la crisis del conductismo en torno a los años 50 (no satisfizo las enormes expectativas creadas en torno a las supuestas leyes universales del aprendizaje, que permitirían a la educación abandonar el reino del arte para adentrarse en la vía de la ciencia) abrió las puertas a uno de los enfoques predominantes hasta nuestros días en psicología: el llamado *cognitivismo*. El objetivo básico de este enfoque ha sido estudiar científicamente lo que había sido dejado de lado a propósito por los conductistas más radicales: la mente humana, aquello que al distinguirnos más y mejor de los demás

animales, nos hacía por ello más específicamente humanos. Los conductistas, al pretender encontrar leyes universalmente válidas, se vieron obligados a prescindir, al menos en un principio, de estas sutilezas diferenciadoras.

La crisis del conductismo conllevó una crisis del modelo clásico en lo que respecta a la concepción del binomio profesor/alumno a la que he hecho referencia anteriormente. El planteamiento conductista de Estímulo-Respuesta o Estímulo-Respuesta-Consecuencia dio paso al de Estímulo-Organismo-Respuesta. Las variables importantes ahora no van a ser los antecedentes o las consecuencias de las respuestas dadas por alumnos y profesores, sino lo que ocurre en sus organismos. Ambos agentes gozan de una mente que les hace considerablemente más complejos de lo que se había pensado hasta esos momentos. Desde el cognitivismo se comenzarán a estudiar una multitud de variables, tanto referidas a los profesores como a los alumnos, que se supone van a jugar un papel relevante a la hora de entender el complejo proceso de enseñanza/aprendizaje. El profesor planifica sus clases de formas muy diversas, tiene sentimientos positivos o negativos con respecto a sus alumnos, asume distintos estilos de enseñanza, cuenta con específicos sistemas de creencias en torno a lo que debe ser la educación, muestra determinadas expectativas de logro que le harán comportarse de formas muy diversas dentro del contexto académico, mostrará diferentes habilidades o destrezas docentes, será portador de variados sistemas de valores, todo lo cual, a su vez, como cabe imaginar va a tener repercusiones positivas o negativas en el otro componente del binomio, el alumnado, en su aprendizaje, y también en el resto del sistema educativo. Algo semejante, pues, ocurrirá con respecto a los alumnos. Ahora ya no aparecerán como seres pasivos que únicamente van a tratar de aprenderse lo que les manden. Se encuentran en etapas evolutivas concretas y muy diferentes entre sí, presentan diversos niveles intelectivos así como determinados grados de desarrollo emocional, ponen en funcionamiento distintas estrategias y estilos de aprendizaje, cuentan con determinados sistemas de valores y creencias condicionados por la sociedad y cultura de la que provienen; todo ello, a su vez, incidirá sobre su grado de aprendizaje y sobre el comportamiento del profesor.

Esta crisis del modelo clásico supuso, igualmente, un cambio sustancial en lo tocante a las relaciones posibles entre profesores y alumnos. Éstas, en la concepción clásica, eran claramente unidireccionales: la sabiduría, los valores, en definitiva toda la enseñanza, debía necesariamente pasar del profesor al alumno. No eran concebibles siquiera otras alternativas. Desde el planteamiento cognitivista, la bidireccionalidad se imponía: el profesor debía contar con la diversidad y las diferencias de todo tipo (intelectivas, afectivas, de valores, culturales) presentes entre sus alumnos y, en consecuencia, hacer todo lo posible por tenerlas en cuenta en su docencia; por su parte, los alumnos debían ser conscientes de la diversidad de modos o estilos de impartir la docencia por los distintos profesores (en función de sus características personales: madurez intelectual y emocional, nivel de conocimientos, motivación) y tratar de acomodarse a esas peculiaridades. La calidad del proceso de enseñanza/aprendizaje en parte, pues, vendrá determinada por los ajustes bidireccionales de alumnos y profesores. El sistema educativo, que como todo sistema es más que la mera suma de las partes, ofrecerá un producto óptimo sólo cuando se cuente con óptimas relaciones bidireccionales. La relación entre ambos agentes educativos, desde esta perspectiva, cobra carta de naturaleza y va a ser incluida como variable relevante, junto con los propios alumnos y profesores, a la hora de explicar la calidad de los diferentes procesos y del producto global de cada sistema educativo.

Todos estos cambios estuvieron acompañados de otro de no menor trascendencia: la *concepción del currículo*, con el que se cierra el primer triángulo (véase Figura 1) de los componentes más básicos de cualquier ámbito educativo: primaria, secundaria o universidad. Frente a la concepción más clásica de currículum cerrado, impuesto desde arriba y definido únicamente por sus contenidos instruccionales, aparece la idea de un currículum abierto (requiere adaptaciones a diversos niveles: nacionales, regionales, de centros, de etapas o ciclos, en cada aula), dialogado entre todas las partes implicadas en el proceso de enseñanza/aprendizaje, aunque con claras y obvias asimetrías, y que englobe, junto con los contenidos a impartir, todas las variables contextuales que rodeen y condicionen dichos contenidos (Lewy, 1991).

Desde este nuevo modelo cognitivista, el profesor ya no aparece como el controlador de todas las contingencias de refuerzos y castigos según las distintas conductas de los estudiantes, sino como el administrador de la información adecuada para que sea procesada activamente por sus alumnos. El símil o metáfora es ahora el ordenador. Sin información (la que ha de proporcionar el profesor, aunque a través de muy diversos medios: clase magistral, lectura de textos en papel o en red, búsqueda por el propio alumno tutelada o guiada) no hay ciertamente aprendizaje posible, pero éste en modo alguno es un proceso pasivo (copia de apuntes) y unidireccional (lo que le imparte de viva voz el profesor). El alumno ha de asimilar lo dado y sugerido por el profesor y lo encontrado por él mismo, reestructurarlo continuamente a medida que se avanza en los distintos niveles educativos, e integrarlo significativamente dentro de su estructura mental: estamos hablando del denominado *aprendizaje significativo* (Ausubel, 1968).

Ya a inicios del siglo XXI van a aparecer algunos *enfoques integradores* que valorando, por una parte, las aportaciones conductistas y, por otra, las derivadas de los distintos enfoques cognitivistas, incorporen asuntos que hasta el momento habían quedado relativamente relegados, debido al sesgo de la sobrevaloración de los contenidos con respecto a los valores. Dentro del enfoque integrador, al que hace referencia la Figura 1, entiendo que alumnos y profesores han de constituirse en una comunidad de aprendizaje/enseñanza, dentro de la cual ambos conjuntamente, aunque por supuesto en diverso grado, van construyendo conocimientos y sistemas de valores que respondan adecuadamente a las necesidades de la sociedad actual globalizada, una de cuyas características más típicas, por oposición a las sociedades pasadas, es el exceso de conocimientos provenientes de las distintas bases de datos, redes de información y medios de comunicación, disponibles en cualquier parte del planeta y a cualquier hora del día y de la noche. En este contexto, parece lógico que la lección magistral que tanto peso ha tenido en los sistemas educativos del pasado, si bien no debiera estar llamada a desaparecer, al menos tendría que ir acompañada de otras muchas técnicas de enseñanza como la exploración a través de los medios informatizados, el descubrimiento guiado, el aprendizaje participativo o el aprendizaje autorregulado (Reigeluth, 1999; Schunk y Zimmerman, 1994).

Este primer triángulo (véase Figura 1) se cierra con las correspondientes flechas bidireccionales entre los currículos (el plural es un indicador claro de los avances típicos del siglo XX a los que he hecho alusión) y los dos agentes protagonistas en cualquier proceso de aprendizaje/enseñanza: profesores y alumnos (cuyo plural tiene semejantes

connotaciones a las de los currículos). Estas relaciones, como ya se indicó, cobrarán tanta importancia a la hora de entender el proceso/producto educativo como la otorgada al alumnado y profesorado, desde una concepción sistémica como la que aquí se defiende.

1.2. Evaluaciones, asesoramientos, intervenciones y contextos

El segundo triángulo está formado por aquellas funciones más básicas que un especialista (un *psicólogo educativo*, por ejemplo) necesariamente ha de ejecutar si asumimos la óptica del continuo perfeccionamiento del propio sistema educativo: las evaluaciones (del profesorado, del alumnado, de las autoridades académicas, de los contextos, de los currículos), los asesoramientos (a profesores, alumnos, padres, máximos responsables académicos) y las intervenciones (con profesores, padres, estudiantes). De ellas, sobre todo de la evaluación, hablaré detenidamente a lo largo de este libro. De nuevo, las flechas bidireccionales nos señalan la estrecha relación que es necesario mantener, por un lado, entre los componentes del primer y segundo triángulo y, por otro, entre los componentes de los dos triángulos y los escenarios en los que ha de desarrollarse cualquier proceso de aprendizaje/enseñanza.

Estos procesos, pues, tienen lugar en contextos muy concretos, que sin duda van a ejercer una influencia determinante en la calidad de todo el sistema educativo (véase Figura 1). Clásicamente se ha hablado de contextos formales (los que dan cabida a los distintos niveles de enseñanza reglada), no formales (aquellos contextos en los que el proceso de enseñanza/aprendizaje se parece al de la enseñanza reglada, pero que no conllevan ningún tipo de habilitación o acreditación oficial) y los informales (el resto de contextos en donde se produce una adquisición de conocimientos, destrezas o valores gracias a la experiencia diaria). A su vez, estos ámbitos pueden ser clasificados como microsistemas (contexto familiar del profesor o del alumno, por ejemplo) mesosistemas (el centro) y macrosistemas (la sociedad). Si se quiere comprender tanto extensivamente como en profundidad cualquier proceso de aprendizaje/enseñanza resulta imprescindible, por tanto, contar con las influencias (armoniosas o contrapuestas) de cada uno de estos contextos en los distintos miembros educativos y las que éstos ejercen en aquéllos (Bronfenbrenner, 1977; Garbarino, 1990).

Con todo ello presente, nos queda por ver cuáles debieran ser los objetivos más generales y a la vez más básicos de los sistemas educativos en sus diversos niveles. Tal como se puede ver en la Figura 1, éstos serían: *la excelencia de procesos/productos académicos, profesionales y personales*. Cuando hablo de excelencia entiendo por tal el logro máximo que cabe obtener teniendo en cuenta las características de todos y cada uno de los componentes, más las pertinentes interacciones bidireccionales hasta ahora señaladas. Esto implica que el hincapié se hace en la competencia no en la competitividad. No se trata de que todos los componentes alcancen la cima de ser el número uno (cosa por otra parte totalmente imposible), sino de que den de sí el máximo que puedan dar (en habilidades, destrezas, conocimientos, actitudes, valores), desarrollados gracias a las interacciones con los demás, en concreto, gracias al aprendizaje mediado de los demás (Vygotsky, 1978).

Al hablar de la excelencia de productos/procesos quiero dar a entender que hoy no es concebible el estudio científico de los unos sin hacer referencia necesariamente a los otros, por lo que pienso que resulta un error querer contraponerlos en tanto objetivos

finales de un determinado sistema educativo. Un análisis mínimo, desde esta perspectiva, tal vez nos ofreciese pistas sumamente valiosas del porqué han fracasado ciertas reformas educativas relativamente recientes en diversos países, al fijarse casi exclusivamente en los procesos y olvidarse de los productos, pues se asumía que los productos eran cosa del pasado.

La excelencia de procesos/productos se centra en dos focos complementarios. Por un lado, parece coherente que si una sociedad invierte en un sistema educativo costoso, en cualquiera de los niveles, quiera obtener unos resultados académicos (buen nivel de conocimientos) y profesionales (buenas habilidades y destrezas) de calidad. Por otro, se constata que si cada uno de los agentes no se ha desarrollado debidamente desde un punto de vista personal (persona equilibrada y satisfecha, individual y socialmente), pueden quedar radicalmente alterados los otros dos objetivos sociales. Como es lógico la materialización de estas metas contará con enormes variaciones en función de las concepciones específicas de las distintas sociedades.

Así, en la mayoría de los sistemas de creencias de los planteamientos educativos vigentes hasta mediados del siglo XX, dominaba a sus anchas el patrón de la unidireccionalidad, al que ya he hecho alusión. Las autoridades, al haber sido constituidas como tales legalmente, ya no tenían que rendir cuentas ni dar explicaciones a nadie. Los profesores, al haber sido seleccionados y acreditados a su vez por las autoridades académicas, no necesitaban ser sometidos a ningún otro juicio de valor. Por el contrario, los alumnos estaban llamados a aprender todo cuanto se les enseñase, por lo que al profesor se le exigía que informara sobre el aprendizaje de cada alumno a fin de que se pudiesen tomar medidas sobre su acceso a niveles superiores de enseñanza o sobre su retención (repetición) o incluso sobre su exclusión del sistema educativo.

En nuestros días este sistema de creencias ya ha mostrado fehacientemente sus debilidades, por lo que las bidireccionalidad y los contextos están cobrando cada día más protagonismo a la par que, como hemos visto, la necesidad de construir compartidamente el conocimiento se va abriendo paso cada vez con más fuerza en nuestra sociedad de la información y del conocimiento. Todo ello implica que la evaluación es requerida para detectar cada fuerza y debilidad de un sistema complejo – el educativo-, capitaneado por unos protagonistas complejos –profesores y alumnos-, que se mueven y se realizan en contextos cada vez más complejos –desde el microsistema familiar, pasando por el mesosistema institucional, hasta llegar al macrosistema social-. Así, no es de extrañar que la llamada *cultura de la evaluación* haya hecho acto de presencia muy a finales del siglo pasado, extendiéndose rápidamente a todos los agentes del ámbito educativo, a los centros de investigación, a las autoridades políticas y a la sociedad en general.

2. Los componentes básicos de la denominada cultura de la evaluación

Los humanos, por el hecho de serlo, estamos condenados a ser permanentes objetos y sujetos de evaluación. No tenemos alternativa. Continuamente estamos evaluando todos y cada uno de los acontecimientos o circunstancias que nos rodean (nuestra mente no puede dejar de hacerlo) y, a su vez, continuamente estamos siendo evaluados por los que nos rodean (“causamos impresiones” –buenas, malas o indiferentes- a todos y cada uno de aquellos con los que interactuamos). Como es lógico esto tiene su plena aplicación en el ámbito educativo. Los profesores se forjan

impresiones (acertadas o no) de sus alumnos y, a su vez, los alumnos tienen sus ideas, con claras connotaciones valorativas, de todos y cada uno de sus profesores. Ambos emiten juicios valorativos sobre el currículo, las familias, las instituciones y los contextos en los que tiene lugar el proceso de aprendizaje/enseñanza. Las autoridades académicas y los padres valoran a los profesores, a los estudiantes/hijos y a todo el sistema educativo. Esto es irremediable. Es lo que se denomina evaluación informal. Desde nuestra aparición como nueva especie, esta evaluación ha acompañado a todos los humanos hasta nuestros días y lo seguirá haciendo hasta su fin. No valen, por tanto, excusas para la “no evaluación”, puesto que, por su esencia constitucional, las personas no podemos dejar de realizarla.

Ahora bien, le ocurre a la evaluación, desde este punto de vista, algo semejante a lo que le sucede a la fuerza de la gravedad. Ésta siempre ha existido y los humanos de todos los tiempos han sido testigos de sus efectos, pero no así la ley de la gravitación universal, que desde esta visión histórica filogenética ha sido establecida recientemente (finales del siglo XVII). De forma parecida, es posible decir que si bien la evaluación informal siempre ha estado presente en el mundo de los humanos, no así la evaluación formal, ya que es un claro producto de la investigación que se inició muy a finales del siglo XIX y que va a tener su apogeo a lo largo de todo el siglo siguiente (Fernández, 2001).

2.1. Etapas de la evaluación

Hoy contamos ya con una breve historia de la evaluación formal, que nos posibilita señalar unas etapas relativamente bien definidas de su progresivo desarrollo científico (Madaus, Scriven y Stufflebeam, 1983). La primera de ellas, la *edad de la reforma*, que abarca todo el siglo XIX, se caracteriza porque nos da cuenta del paso de las evaluaciones informales (escasamente científicas) a la evaluación formal, en tanto ésta conlleva la aplicación de una metodología experimental que tan buenos resultados había producido dentro de las denominadas ciencias duras (Rice, 1897, 1914). Los elementos básicos son: grupo experimental, grupo control, programa de intervención o reforma y, como colofón, la evaluación final que había sido precedida de la evaluación inicial. La presencia o ausencia de diferencias estadísticamente significativas (experimento más estadística) será la base, en la evaluación formal –científica-, del juicio de valor, frente a la mera impresión, que es la base típica de las evaluaciones informales.

La segunda etapa, la *edad de la eficacia o de los tests*, abarcará aproximadamente el primer cuarto del siglo XX y se va a caracterizar tanto por la elaboración de todo tipo de instrumentos de evaluación con los que garantizar, de forma válida y fiable, la recogida de la información como por la obsesión por la eficacia: se pretendía saber rigurosamente el grado de consecución de los objetivos establecidos en cualquier *acto propositivo*, en nuestro caso de tipo educativo (aprendizaje de los alumnos, enseñanza del profesorado, implementación del currículo, etcétera).

El tercer periodo, la *etapa tyleriana*, denominada así en honor a la figura evaluadora más destacada del segundo cuarto de la pasada centuria (Tyler), va a suponer un avance considerable en casi todos los terrenos de la evaluación formal. Es Tyler quien va a acuñar el término de *evaluación educativa* y quien, sobre todo, va a proporcionarnos un enfoque concreto que todavía se sigue utilizando en nuestros días. Su concepción queda perfectamente reflejada en su definición de la evaluación en tanto valoración del grado en

que unos objetivos propuestos han sido logrados, debido a la implantación de un determinado programa (Tyler, 1942, 1950, 1967). Este enfoque supone un importante cambio de orientación con respecto al enfoque de tipo experimentalista de Rice, con el que se pretendió que la evaluación entrara por primera vez en la vía de la ciencia. El método experimental va a ser sustituido por el método comparativo, gracias al cual la evaluación formal tendrá un único y claro objetivo: poner de manifiesto, de forma fiable y válida, el grado en que los objetivos establecidos *a priori* de cualquier acto propositivo (curso, programa, asignatura,...) fueron cumplidos.

En torno a los años 50 del mismo siglo se desarrollará el cuarto periodo, denominado *etapa de la inocencia o de la irresponsabilidad social*. En estos años asistimos ciertamente a un considerable avance, tanto teórico como tecnológico, de los más diversos instrumentos de evaluación, aunque estos desarrollos científicos no se transformaron en beneficios tangibles de carácter social. De ahí que la propia denominación de este periodo ya nos advierta de la escasez de logros educativos debidos a la aplicación de la evaluación formal.

La siguiente década, que dará origen a un nuevo periodo, la *etapa del realismo o de la expansión*, estará claramente condicionada por las críticas a las deficiencias de la anterior. De ahí que se intente dar ante todo respuesta a la irresponsabilidad social mediante la puesta en funcionamiento de los distintos enfoques desarrollados hasta entonces: el riceano, el psicométrico, o el tyleriano, fundamentalmente. En pleno apogeo de esta práctica evaluadora se va a producir, curiosamente, una grave crisis de la evaluación formal: su práctica masiva hizo aflorar la enorme distancia existente entre las grandes expectativas puestas en las potencialidades de la evaluación científica para distinguir lo eficaz de lo ineficaz (por ejemplo, en los distintos sistemas de enseñanza y aprendizaje) y los resultados obtenidos. Ante este duro revés, los evaluadores se vieron en la necesidad imperiosa de intentar ofrecer algún tipo de solución que evitara el incipiente descrédito de la evaluación. Se pensó en la conveniencia o necesidad de una nueva concepción de la evaluación o en el establecimiento de nuevos enfoques, como algunas de las posibles alternativas.

En los años 70 tendrá lugar un nuevo periodo, denominado *etapa del profesionalismo*, que se inicia, como cabía esperar a raíz de lo acaecido en la etapa anterior, con una crisis de identidad. En este quinto periodo nos hallamos hoy en día y el objetivo más importante está siendo el de dar una solución científica a tal crisis, es decir, encontrar cuáles debieran ser las funciones básicas a desempeñar por parte del evaluador de forma que, al menos en parte, las grandes expectativas en torno a la evaluación de los distintos actos propositivos (proyectos, programas, cursos, currículos, sistemas de enseñanza/aprendizaje, etcétera) puedan ser razonablemente satisfechas, es decir, saber con cierto rigor si son eficientes o ineficientes, más allá de las primeras impresiones sobre las que se fundamentaban las evaluaciones informales.

Es pertinente indicar que si bien son varios los autores que refrendan estas etapas (Stufflebeam y Shinkfield, 1985), en modo alguno queda totalmente zanjada esta cuestión histórica, dado que de hecho han surgido otras alternativas –la de las distintas *generaciones* de evaluaciones: estaríamos ahora en la cuarta- a la hora de concebir esta evolución de la evaluación formal a lo largo del siglo XX (Guba y Lincoln, 1989). Estos autores establecen sólo cuatro etapas o generaciones: la de la medición, la de la descripción, la del juicio o valoración y la constructivista, considerada esta última como la

visión más acorde con los conocimientos y planteamientos evaluativos de finales de siglo. El análisis complementario de estos dos enfoques de la historia puede, sin duda, enriquecer la comprensión actual de la evaluación formal, en un sentido muy general. Si el objetivo que uno tiene en mente, por el contrario, es contar con una visión histórica más específica de la evaluación formal, centrada en el ámbito de la educación, la puede encontrar en Worthen y Sanders (1987), quienes sólo establecen tres grandes etapas: la anterior a 1920, la que transcurre desde 1920 hasta 1965 y la que va desde esta última fecha hasta nuestros días.

2.2. Institucionalización de la evaluación

La crisis de la evaluación formal, surgida al inicio del último cuarto del siglo pasado, va a encontrar varias vías complementarias de solución, la mayoría de ellas cubiertas bajo el paraguas de la *institucionalización de la evaluación*. Éste es un remedio ya bien conocido, debido al éxito manifiesto en otras disciplinas plenamente consolidadas. Se destacarán aquí sólo algunas de esas *vías* más significativas.

Qué duda cabe de que no es posible que un campo se consolide científicamente si no aparecen publicaciones sobre la materia en medios –revistas, tanto de carácter general como específicas- que ofrezcan una garantía de lo dado a la luz pública. El sistema de control de calidad predominante en la comunidad científica internacional es, en estos momentos, el conocido como revisión por pares. Un nivel de control superior de la calidad viene definido por el índice de impacto de la revista (razón entre citas obtenidas y número de artículos publicados durante un periodo de tiempo determinado). Dentro del ámbito de la evaluación, ya se cuenta en la actualidad con un buen número de revistas que satisfacen el criterio más exigente (su inclusión en Journal Citation Reports –índice de impacto-): *American Journal of Evaluation, Evaluation and Program Planning, Evaluation Review, Research Evaluation,...* Si acotamos más el campo y nos detenemos en la evaluación educativa, son bien conocidas: *Assessment and Evaluation in Higher Education, Educational Evaluation and Policy Analysis, Educational Research and Evaluation, Journal of Personnel Evaluation in Education, Studies in Educational Evaluation, ...* Si ahora ampliamos el campo de miras para hacernos una idea más global de lo existente, nos encontramos con más de 800 revistas, dentro de la base de datos ISSN, que incluyen en su título la palabra evaluación (*Evaluation: The International Journal of Theory, Research and Practice, Evaluation Practice, Evaluation Quarterly, Language Resources and Evaluation, Measurement and Evaluation in Counseling and Development,...*).

A esto ha de añadirse el saber más consolidado, al no tener el carácter de inmediatez típico de las revistas, que aparece en series, manuales, enciclopedias, diccionarios, libros básicos o revisiones periódicas, que debieran formar parte del canon del buen evaluador. Así, en el caso de las series, destacaremos la más conocida internacionalmente: “Evaluation Studies Review Annual” (sirva el volumen de Shadish y Reichardt, 1987, como un ejemplo de toda la serie). En cuanto a manuales, citaremos entre otros, el de Bloom, Hastings y Madaus (1971), el de Chelimsky y Shadish (1997), el de Herman, Morris y Fitz-Gibbon (1987), el de Kellaghan y Stufflebeam (2003), el de Millman (1981) o su nueva versión/edición de Millman y Darling-Hammond (1990), el de Struening y Guttentag (1975) y el de Wholey, Hatry y Newcomer (1994). En el terreno de las enciclopedias, haremos referencia, entre otras, a la de Mathison (2004) y a la de Walberg y Haertel (1990). El campo de la evaluación se ha extendido tanto entre las más

diversas disciplinas que se ha hecho incluso necesaria la aparición de un diccionario destinado a la clarificación conceptual (Landsheere, 1985).

Por lo que respecta a los libros básicos baste decir que su sola enumeración ocuparía varias páginas. Por eso aquí sólo haré mención al conjunto de 9 volúmenes (Program Evaluation Kit) por ofrecer, en su conjunto, una visión muy completa sobre los pasos básicos y fundamentales en cualquier proceso de evaluación: el diseño, el análisis de los datos o la presentación y comunicación de los resultados (se puede ver, como ejemplo, el volumen de Fitz-Gibbon y Morris, 1987, sobre cómo diseñar una evaluación de programas). Ni que decir tiene que ya se cuentan por miles, como se comprueba mediante una visita rápida a la base de datos ISBN, los libros en cuyos títulos aparece la palabra evaluación o alguno de los sinónimos.

De entre las “revisiones descriptivas” periódicas, circunscritas al ámbito de la psicología, hay que destacar sin duda las aparecidas en la revista *Annual Review of Psychology*. Su consulta resulta imprescindible para el buen evaluador, pues a través de las mismas es posible seguir los avances y aportaciones más importantes en el terreno de la evaluación durante un determinado periodo de tiempo, en este caso, las últimas décadas del siglo pasado (Cook y Shadish, 1986; Glass y Ellet, 1980; Lipsey y Cordray, 2000; Perloff, Perloff y Sussna, 1976; Sechrest y Figueredo, 1993; Wortman, 1983). Otros artículos también resultan de especial significación, para el especialista y para el que se inicia en la materia, pues tratan de proporcionarnos una visión sintética del campo analizado: sus mitos y realidades, sus autores y trabajos más relevantes, sus conclusiones mejor fundamentadas, etcétera (Abrami, D’Apollonia y Rosenfield, 1996; Aleamoni, 1999; Darling-Hammond, Wise y Pease, 1983; Gordon, 1998; Marsh, 1987; Nevo, 1983; Trevisan, 2004).

A todo ello se podrían añadir los cientos de capítulos de libros que aparecen en obras de muy diversas materias y que están dedicados a poner de manifiesto las evaluaciones concretas en esas áreas específicas de conocimiento (pueden verse, como ejemplos, los manuales de Bickman y Rog (1998), de Lindzey y Aronson (1985) o de Reason y Bradbury (2001).

Además de los conocimientos científicos, bien hayan aparecido éstos en revistas, libros o capítulos de libros, se ha hecho necesaria la creación de asociaciones e instituciones encargadas y dedicadas en exclusividad a la consolidación de un campo científico, en este caso el de la evaluación formal. Baste señalar aquí para ratificar lo dicho al *Grupo 12 de Mayo*, la *División H de la AERA* o las distintas asociaciones surgidas en diferentes países como la *Sociedad Alemana de Evaluación*, la *Sociedad Americana de Evaluación*, la *Sociedad Británica de Evaluación*, la *Sociedad Canadiense de Evaluación*, la *Sociedad Europea de Evaluación* o la *Sociedad de Evaluación de Australasia*, entre otras. Gracias a las páginas electrónicas de estas asociaciones es posible obtener información de sus publicaciones, actividades y enlaces informáticos con otros sitios especialmente relevantes para la evaluación.

A todo ello hay que añadir que se han creado centros dedicados específicamente a la investigación y desarrollo de la evaluación (Center for Instructional Research on Curriculum Evaluation, The Stanford Evaluation Consortium, entre otros muchos), además de los múltiples departamentos dentro de distintas universidades de diferentes países que hacen desaconsejable siquiera su mera enumeración. Igualmente, merece la pena hacer mención de los diversos comités encargados de establecer unas normas consensuadas para

llevar a cabo evaluaciones formales (Evaluation Research Society –ERS- Standards Committee o Joint Committee on Standards for Educational Evaluation). De hecho, el volumen que aparece a comienzos de la década de los 80 publicado por este último Comité sobre normas de evaluación para programas, proyectos y material educativo, bien en esta primera edición o bien en la segunda, se constituye en obra de referencia obligada para cualquier buen evaluador que desee trabajar en el ámbito educativo (JCSEE, 1981; 1994). Algo semejante cabría decir de los otros dos libros publicados a finales de los 80 (JCSEE, 1988) y comienzos de la presente década (JCSEE, 2003), respectivamente, que complementan a los dos citados, al centrarse ahora en la evaluación de los miembros que participan en los contextos de enseñanza y aprendizaje.

2.3. Funciones del evaluador

Dada la abundancia de información con la que contamos (a todas luces inabarcable para cualquier ser humano aun cuando le dedicase toda su vida a la lectura de lo hasta este momento publicado), ¿podríamos al menos saber *grosso modo* cuáles son consideradas las principales funciones a desempeñar por el evaluador de nuestros días, con el fin de lograr alguna salida a la crisis de identidad con la que se iniciaba la última etapa, la del profesionalismo?

En lo que coinciden casi todos los expertos evaluadores es en que la respuesta a esta pregunta va a depender necesariamente de la perspectiva inicial en la que uno decida instalarse. En un intento de síntesis, se han destacado al menos cinco (Stecher y Davis, 1987). Así, los que entienden que investigación y evaluación tienen un denominador y objetivo común, siendo la evaluación la versión aplicada de la investigación básica, reclamarán un *evaluador experimentador/científico*, es decir, aquella persona capaz de controlar variables dependientes e independientes, gracias a lo cual pueda demostrar relaciones auténticas de causalidad lineal. Ciertos comportamientos del profesor podrían ser determinantes (variables independientes) a la hora de entender cierto grado de rendimiento del alumno (variable dependiente), tal como se ha venido proponiendo desde uno de los enfoques más clásicos en investigación educativa: el conocido como modelo instruccional proceso-producto. El énfasis se pone en el diseño experimental o cuasi-experimental, en el control de las variables. Su debilidad tal vez resida precisamente en la dificultad o imposibilidad de poder llevar a cabo tal control en contextos tan sumamente complejos como el que hemos descrito del ámbito educativo. Cuando éstos se simplifican a fin de lograr el control, surge entonces el problema en torno a la generalización de las conclusiones a otras situaciones distintas de aquella en la que tuvo lugar el experimento, es decir, el problema de la validez ecológica de los resultados.

Los que adoptan una perspectiva típicamente tyleriana o conductista, orientada a metas u objetivos, piensan que el evaluador ideal sería una *persona especializada en la medida*, dado que su función básica es determinar, de forma válida y fiable, hasta qué punto los objetivos propuestos en un determinado acto propositivo (programa o proyecto) son logrados. Lo más importante aquí es, por un lado, especificar con claridad y objetividad los objetivos y, por el otro, acertar con la técnica de medida más adecuada. Si la fortaleza de esta perspectiva reside en los esfuerzos realizados para ser claros en torno a las metas anheladas y rigurosos en la recogida de datos, su debilidad apunta a la visión de túnel que a veces tiene lugar cuando todas las miradas van dirigidas hacia unas determinadas dianas, produciéndose una cierta ceguera con respecto a los fenómenos o consecuencias no previstos *a priori*. Por lo que atañe a la medida (recogida de datos), no

siempre se da razón de ser de la compatibilidad entre los fenómenos humanos analizados y los requisitos (presupuestos) de las técnicas empleadas.

Desde la perspectiva de la toma de decisiones, se requiere un evaluador que sea un *especialista en toma de decisiones* para que pueda avalar y apoyar a los gestores (académicos, políticos, empresariales) que hayan de ejecutarlas. La preocupación fundamental del evaluador ha de ser la de reunir información de forma sistemática y contrastada en torno a aquellos asuntos de interés sobre los que sea necesario un pronunciamiento y su correspondiente puesta en funcionamiento. Su punto fuerte reside en la amplitud de miras (distintas fuentes de información y distintas técnicas para su recogida) y tal vez su mayor debilidad provenga de las dificultades a la hora de discriminar entre información relevante frente a irrelevante. A veces los condicionamientos de los plazos o presupuestos hacen que la calidad de la información no sea precisamente la más adecuada para tomar decisiones bien fundamentadas.

Desde una cuarta perspectiva, la orientada al usuario, se concibe al evaluador como un *colaborador*. Su función principal será la de tratar de entender cuáles son las necesidades reales de los usuarios, a fin de que éstas puedan ser correctamente satisfechas. Su punto fuerte se manifiesta en la cercanía a los usuarios a los que trata de servir, proporcionándoles las informaciones sin sesgos que realmente necesitan, para estar acertados en sus decisiones vitales. Su lado más débil aparece cuando otros intereses más poderosos que los de los usuarios hacen acto de presencia, introduciendo un sesgo indebido en las informaciones proporcionadas a los usuarios.

Desde la perspectiva denominada “sensible”, se reclama a un evaluador que sea sobre todo un *asesor*. Su labor esencial es procurar comprender todos los puntos de vista existentes ante un determinado acto propositivo (un currículo, por ejemplo), con el fin de tratar, en primer lugar, de tenerlos en cuenta y, después, de satisfacerlos en la medida de lo posible (puntos de vista de los profesores, de los padres, de los alumnos o de las autoridades políticas o académicas dentro de un determinado ámbito o nivel educativo). Su principal punto fuerte se manifiesta en la consideración de todas las visiones de los interesados, mientras que su debilidad aparece cuando no se cuenta con un sistema de prioridades o jerarquías que haga viable cualquier toma de decisiones sensata y bien argumentada.

A la luz de lo expuesto, cabe inferir que si bien no hay un total consenso sobre las funciones del evaluador, pues éstas dependen de la perspectiva desde la que se parte, sí parece que se va produciendo un avance de síntesis en torno a las mismas, de forma que a medida que progresamos en conocimientos sobre la evaluación formal, es posible ir dando a luz un perfil de competencias básicas y competencias específicas del buen evaluador (King y Stevahn, 2002; King, Stevahn, Ghery y Minnema, 2001).

2.4. Plaga o remedio para la mejora de lo evaluado

Hasta aquí se ha expuesto una panorámica de la institucionalización de la evaluación formal desde arriba, es decir, desde los avances científicos y desde la aceptación de la misma por los distintos tipos de autoridades: académicas, políticas o administrativas. Pero, ¿qué ocurre con la institucionalización desde abajo, es decir, desde las perspectivas de los afectados, desde las opiniones de los que son evaluados? En algunas ocasiones, ¿no resulta como una especie de moda/plaga que todo lo invade, que es

pasajera, que no sirve prácticamente para nada salvo para cumplir, al menos aparentemente, con la obligación de rendir cuentas que tienen las distintas autoridades ante los contribuyentes? No cabe duda de que muchas veces los evaluados tienen razón. Es inimaginable la cantidad de informes de evaluación que duermen el sueño de los justos (injustamente, por cierto, en buena parte de los casos), después de haberles prometido a los evaluados que sería muy útil para su mejora profesional. De ahí el comentario al que frecuentemente ha de enfrentarse el evaluador: díganos de verdad si en esta ocasión va a servir para algo la evaluación, ya que la experiencia indica lo contrario. Junto a esa actitud, el evaluador también puede encontrarse con un tipo de evaluado suspicaz que considera que la evaluación no es más que un mecanismo/truco del que va a servirse la autoridad para tomar decisiones administrativas no del todo justificadas.

Ante esta situación no tan infrecuente, se requiere dar unos primeros pasos que faciliten y allanen el camino de la evaluación formal, pues ésta debiera ser algo más que una mera cuestión técnica, como han puesto de manifiesto McLaughlin y Pfeifer (1988). Tras la experiencia acumulada en varios centros, estos autores nos indican que antes de implantar un proceso evaluador se requiere preparar a la institución para un cambio sustancial dentro del cual tenga cabida una doble modificación de actitudes. Por un lado, la negativa (de resistencia o incluso fobia) de los evaluados y, por otro, la positiva: el compromiso serio por parte de las autoridades de proporcionar la ayuda adecuada para el perfeccionamiento de los docentes y, por tanto, para la mejora del proceso/producto educativo, gracias a los datos obtenidos en la evaluación formal. Ello exige crear antes que nada un clima propicio para la evaluación, que conlleve un cambio de valores y de estrategias conductuales. Es necesario que las autoridades, decanos o directores de centros, se ganen la confianza de los evaluados, rompiendo el círculo de la distancia (se han de compartir sistemas de valores, expectativas, metas y entrenamiento práctico en todo lo referido a la evaluación). No se les ha de confundir, o lo que es peor engañar, tratando de implantar una evaluación meramente cosmética o ritualista (para cubrir las apariencias y salvar la cara de la dirección) o con una finalidad distinta de la anunciada (tomar decisiones administrativas sobre su contrato cuando se comenzó hablando de ayuda para alcanzar mayores niveles de excelencia docente). Las autoridades han de ir comunicando los distintos pasos del proceso evaluador a los evaluados, a la par que compartiendo los problemas y los logros, de modo que se establezca una bidireccionalidad equilibrada en vez de una unidireccionalidad manifiestamente asimétrica (la dirección se reserva el grueso de la información relevante), pues sólo de esa forma es posible acrecentar la confianza mutua, que resulta imprescindible para que se pueda implementar con éxito cualquier tipo de evaluación.

Para que este cambio institucional ocurra, a veces resulta imprescindible que aparezca un evento desencadenador (cambio de dirección, por ejemplo), a partir del cual surja un liderazgo que garantice los imprescindibles recursos humanos y económicos, que se gane la confianza de los distintos agentes evaluados y que los implique en todos y cada uno de los momentos importantes del proceso evaluador. El evaluado siente así que tiene voz y voto, que se le consulta, que puede participar en el diseño de la evaluación, que sabe en todo momento la finalidad o finalidades de la misma, que puede disentir de ciertos planteamientos o manifestar sus dudas con respecto a determinados instrumentos de recogida de la información, que puede compartir dudas, temores y también gozos con sus compañeros y con la dirección. Ha de poderse sentir coautor de la propia evaluación, por lo que también participará en las decisiones ulteriores, bien estén encaminadas a la mejora de la enseñanza (corrigiendo defectos o perfeccionando estrategias) o bien a rendir cuentas

(de competencias o incompetencias) ante los contribuyentes. No cabe la menor duda de que todo ello lleva su tiempo y que además es un proceso sin fin, pues cuando esta cultura compartida de evaluación se quiebra, la evaluación se hace imposible.

3. Comentarios

A la luz de lo expuesto, parece claro que el enfoque o modelo clásico de la enseñanza (el profesor como mero transmisor de conocimientos) ha de ceder su puesto a un nuevo *modelo integrador*, dentro del cual se considere a cada miembro o componente en toda su complejidad. Hoy contamos con datos rigurosos que avalan dicha complejidad (sabemos más de las etapas evolutivas de los alumnos, de su diversidad intelectual y cultural, de sus diversos grados de motivación, al igual que conocemos más sobre los distintos estilos de enseñanza de los profesores, de sus diferentes destrezas pedagógicas o de sus distintos grados motivacionales). Además, las relaciones que mantienen entre sí (bidireccionales más que unidireccionales, aunque sin dejar de ser asimétricas) han de ser consideradas como entidades en sí mismas, pues se constituyen en factor clave a la hora de poder explicar la calidad de un proceso/producto educativo. Dentro del nuevo modelo cobra especial importancia la atención a los contextos (microsistemas, mesosistemas y macrosistemas) en donde tiene lugar cualquier proceso de aprendizaje/enseñanza. Las metas, como es lógico, también varían considerablemente de un modelo a otro. El objetivo fundamental no puede ser, sobre todo en enseñanza secundaria y universitaria, la mera transmisión de unos conocimientos que hoy se pueden encontrar, con los medios informáticos adecuados, por doquier. Se impone como meta el enseñar a pensar o el enseñar a aprender (Nickerson, Perkins y Smith, 1985). Además, alumnos y profesores han de convertirse ante todo y sobre todo en permanentes aprendices, por exigencias sobre todo del continuo avance de los conocimientos científicos que es consustancial a nuestra sociedad. Ambos están llamados a conocer lo existente (transmisión de la herencia cultural), pero sobre todo a hacer avanzar más los conocimientos. Además, ambos han de contribuir a la expansión de un sistema de creencias y valores (los propios de los sistemas democráticos) y a un cambio de actitudes (lucha contra la discriminación en función de edad, sexo, nación o etnia, a la par que valoración de los comportamientos cívicos: colaboración, apoyo y ayuda).

Ante unas metas como las propuestas se hace imprescindible contar con la *evaluación formal* más que con la valoración informal, pues estamos dentro de contextos académicos. Sólo ella nos permitirá ir viendo el grado de consecución de las mismas, a la vez que nos posibilitará reflexionar sobre los pasos a seguir para su mejor consecución. La evaluación formal cuenta ya con una historia y una institucionalización que garantiza su utilización científica, aunque es necesario saber que no contamos con un modelo único de evaluación, por lo que las funciones del evaluador pueden ser bastante distintas en función de la perspectiva adoptada *a priori*. En todo caso, la implementación de cualquier modelo evaluativo requiere que previamente se lleve cabo un cambio institucional en sus sistemas de creencias y valores, a fin de contrarrestar, por un lado, las actitudes negativas de los evaluados y, por otro, crear un clima de confianza y colaboración a lo largo de todo el proceso evaluador en cuanto tal (Taut y Alkin, 2003). Este clima ha de fundamentarse en el hecho incuestionable de que todos (autoridades, profesores, alumnos, padres) somos sin excepción, por el hecho de ser humanos, permanentes sujetos y objetos de evaluación. Aquí ha de hundir sus raíces la *cultura de la evaluación*, tan de moda en nuestros días, a juzgar por su abundante presencia en los medios político-académicos nacionales e

internacionales e incluso por su aparición en publicaciones científicas que han visto la luz tanto en formato tradicional como electrónico (Kells, 1995; Scheerens, 2004).

CAPÍTULO SEGUNDO

Evaluabilidad, evaluación y metaevaluación

En el primer capítulo pudimos comprobar cómo la *evaluación formal* hizo su entrada en la vía de la ciencia a finales del siglo XIX, cómo se ha ido desarrollando científicamente a lo largo de todo el siglo XX y cómo al inicio de esta centuria la encontramos ya totalmente institucionalizada. Incluso es frecuente leer y oír hablar de la *cultura de la evaluación* como algo consustancial a nuestra sociedad, al exigir ésta cada vez más unos niveles de perfeccionamiento superiores en todos los terrenos (en nuestro caso: mejores profesores, mejores alumnos, mejores procesos/productos) y también al tener que rendir cuentas todas las instituciones de los presupuestos a ellas asignados, bien por autoridades públicas o privadas.

En el presente capítulo voy a tratar de aclarar lo más posible los tres momentos imprescindibles de cualquier proceso de evaluación formal, tal cual se puede ver en el Cuadro 1: las valoraciones de la evaluabilidad, los distintos tipos de evaluación y sus distintos enfoques y, por fin, las metaevaluaciones (Fernández, 2001).

Cuadro 1
Los tres momentos básicos del proceso evaluador

VALORACIONES DE LA EVALUABILIDAD	EVALUACIONES	INICIALES	META-EVALUACIONES
		DE PROCESOS	
		DE PRODUCTOS	

1. Valoraciones de la evaluabilidad

Posiblemente estemos ante uno de los conceptos menos conocidos incluso por los propios evaluadores, mucho menos por los responsables políticos y resulta casi un perfecto desconocido para un público por lo demás cultivado, pues me estoy refiriendo al mundo universitario. Habrá que empezar, consecuentemente, por un primer acercamiento a su posible definición. Se suele considerar la valoración de la evaluabilidad como un proceso sistemático encaminado a determinar la viabilidad de una evaluación, es decir, a planificar evaluaciones útiles. Este proceso nos ha de ayudar a saber si los resultados de la evaluación nos servirán para mejorar lo evaluado (en nuestro caso: profesorado, alumnado, currículo, autoridades académicas, institución educativa). Nos debiera ayudar también a identificar factores o circunstancias que muy probablemente, si ocurren, imposibiliten o dificulten la evaluación.

Tal vez sea Rutman (1980, 1984) uno de los autores que más ha intentado aclarar la importancia de este concepto. No en vano le ha dedicado todo un libro a esta cuestión. Intentaré seguir en este breve apartado algunos de sus pasos. Para él, la valoración de la evaluabilidad equivaldría a una preevaluación de la evaluación, gracias a la cual sería posible determinar la viabilidad de esta última. Dado que toda evaluación se ha de centrar en un acto propositivo (programa, proyecto, enseñanza, aprendizaje), la valoración de la evaluabilidad posará su mirada en ese acto con el fin de analizar pormenorizadamente, en primer lugar, las características esenciales del mismo: si está bien definido, si las metas y los efectos están claramente especificados y si están bien establecidas las relaciones correlacionales o causales de las variables. Tengo por hábito no

comenzar ni una sola evaluación hasta que previamente los responsables académicos, empresariales o políticos no me hayan proporcionado un escrito en el que se dé cuenta detallada de todo lo anterior. En este contexto, la finalidad de la valoración de la evaluabilidad haría referencia al cuestionamiento previo de todas aquellas características del programa, que aun siendo sumamente atractivas para los posibles usuarios -que los profesores sean competentes, que muestren una pedagogía docente excelente, que estimulen satisfactoriamente el aprendizaje creativo de los alumnos; que éstos a su vez sean creativos, críticos, que razonen con corrección lógica, que sean capaces de expresarse oralmente y por escrito en uno o varios idiomas- distan mucho de gozar de la suficiente concreción como para que puedan ser evaluadas. Se necesita transformar, por tanto, el “lenguaje político” (metas retóricas, por ejemplo), cargado generalmente de connotaciones positivas (mejora de la enseñanza, excelencia docente, investigación de calidad, aprendizaje creativo) a fin de que los electores les den sus votos, en un lenguaje mucho más concreto, gracias al cual se pueda proceder a la realización de la evaluación. Cada idea o concepto ha de ir acompañado de las pertinentes concreciones. Si se habla, imaginemos, de creatividad, la pregunta inmediata es ¿me podría poner algunos ejemplos concretos gracias a los cuales todos pudieran ver lo que realmente se pretende con ese acto propositivo llamado a fomentarla?

Seguidamente, en segundo lugar, se pasará revista a la viabilidad de su implantación o, más específicamente, de su implementación. ¿Existe una cultura de la evaluación que garantice la correcta puesta en funcionamiento de dicha evaluación? ¿Se cuenta con el personal técnico adecuado? ¿Hay seguridad de que se podrá disponer de los suficientes recursos materiales y humanos? ¿Existe una voluntad firme de sacar provecho de los datos de la evaluación? ¿Se cuenta con los instrumentos de valoración más adecuados? ¿Qué relación cabe establecer entre los costes y los beneficios? ¿Se cumplen los protocolos éticos y legales vigentes en esos momentos? ¿Hay coincidencia entre propósitos manifiestos y encubiertos? ¿Vale la pena iniciar una evaluación de unos objetivos insustanciales o carentes de valor real?

Para poder responder satisfactoriamente a estas preguntas básicas se requiere que, contrariamente a los hábitos actuales, diseñadores de programas de intervención, políticos y evaluadores trabajen conjuntamente, de forma que antes de poner en funcionamiento cualquier programa se pueda asegurar con cierto rigor -este es el oficio de la valoración de la evaluabilidad- que el programa, una vez implementado, pueda, posteriormente, ser evaluado. La cantidad de tiempo y de dinero que puede ahorrar una valoración de la evaluabilidad correctamente llevada a cabo salta a la vista. Por desgracia, la cruda realidad nos indica que este primer paso casi nunca se lleva a cabo con seriedad. De hecho abundan las evaluaciones que Rutman califica como de “caja negra”, entendiendo por tal las evaluaciones que son ejecutadas sin haber sido especificadas las metas, los medios de la implementación, las implicaciones o su posible utilización. Tal vez haya llegado el momento, a comienzos de este siglo XXI, de que pueda ser tomada en serio la valoración de la evaluabilidad por los que trabajan en el extenso campo de la evaluación y por los responsables políticos y académicos.

Toda valoración de la evaluabilidad conlleva, generalmente, considerables modificaciones de las propuestas iniciales. Los actos propositivos suelen acabar estando mejor estructurados; las metas mejor especificadas, más realistas; los diseños más elaborados; las técnicas mejor seleccionadas; los análisis de datos más ajustados; los imaginables conflictos entre los distintos participantes prácticamente resueltos; los posibles

efectos secundarios detectados; los problemas éticos o legales resueltos. Dos factores esenciales han de estar presentes para que esto suceda:

- que se cuente con un *profesional de la evaluación experimentado* (sin él las probabilidades de fracaso son muy altas por muy buenas intenciones que se tengan, pues estamos hablando de una evaluación formal);
- que se cree ante todo una *cultura de la evaluación*, gracias a la cual tenga lugar un clima de confianza entre todos los miembros que van a participar en el proceso evaluador.

2. Las evaluaciones

Por lo que respecta al segundo paso -la evaluación- no contamos con una definición universalmente consensuada. Es más, como cabe imaginar, existen decenas de ellas. De ahí que nos tengamos que ir acostumbrando a los plurales como algo consustancial a las ciencias sociales. Ante esta situación, algunos autores (sirvan de ejemplo Glass y Ellett, 1980) optan por ofrecernos una muestra ilustrativa de definiciones a fin de que cada lector pueda hacerse de esta manera una idea mínimamente cabal del área de la evaluación. Otros, como es mi caso (no muy frecuente, por cierto), prefieren hacer un barrido del máximo número de definiciones posibles, después someterlas a un proceso de minucioso escrutinio y, finalmente, ofrecer al lector el resultado final de ambos procesos. Paso así a mostrar, a continuación, lo que entiendo por características esenciales o denominador común de un amplio espectro de definiciones.

Creo que desde la mayoría, si no todas, las concepciones de evaluación se asume que ésta consiste en una *recogida sistemática de la información, con su correspondiente interpretación, en torno a un acto propositivo*. La rica variación surge cuando se contempla dicha recogida mediante los más diversos instrumentos: auto y hetero-observación (con sus diversas clases), entrevistas (también con sus múltiples formas), cuestionarios (se cuentan por cientos), encuestas, grabaciones, etcétera.

Con esta recogida plural de la información se pretende, al menos en la mayoría de los casos, que ésta pueda servir de fundamento sólido para las correspondientes *decisiones*. A partir de aquí, se abren dos mundos bien diferentes. Por un lado, están los evaluadores que entienden que las decisiones han de tener ante todo y sobre todo un *carácter formativo*, es decir, siempre encaminadas a la mejora de lo evaluado, mientras que, por el otro, nos encontramos con los evaluadores que destacan el *carácter sumativo* de tales decisiones, es decir, piensan en actos administrativos con importantes implicaciones para lo evaluado (su continuación, su mejora o su paralización). Ahora bien, esta doble perspectiva no pone en tela de juicio en modo alguno el primer aserto -el del apoyo empírico para cada toma de decisión-, válido para ambos tipos de funciones.

La recogida de la información mediante diversas técnicas que nos garanticen la fiabilidad y la validez de la misma servirá, además de para poder tomar decisiones bien fundamentadas, para *rendir cuentas* ante quien posibilitó que se llevara a efecto un determinado acto propositivo, bien con fondos públicos o privados. El rendir cuentas puede entenderse y representarse como un continuo. Éste puede ir desde la concepción más economicista posible, con sus técnicas de análisis de costes/beneficios (tanto se ha invertido, tanto se ha obtenido de beneficios), asumiendo que los beneficios han de ser

superiores siempre a los costes, hasta la más “humanista”, ética y democrática, cuyo énfasis recae en la responsabilidad ante los demás, sabiendo que aquí los costes materiales siempre han de estar totalmente subordinados a los beneficios humanos, dado que estos últimos son considerados fines a los cuales se han de subordinar los medios económicos. Se pueden dar, y de hecho se dan, posturas intermedias en las que las entradas suelen poderse traducir a la postre en variables de tipo económico (lo que cuesta materialmente un programa de perfeccionamiento del profesorado, por ejemplo), pero las salidas se materializan mediante variables que implican unas mejoras humanas y sociales (superación del fracaso escolar, mejora de la calidad de la enseñanza, desarrollo profesional del profesorado), de forma que los beneficios se sustituyen ahora por la efectividad o la eficacia en el logro de los objetivos propuestos.

Casi todos los enfoques de evaluación tienen como meta la detección del grado de *eficacia/logro* (relación entre las metas propuestas y los resultados) o de *eficiencia* (relación entre las entradas -recursos humanos o materiales, el tiempo invertido- y las salidas -los resultados-) de lo evaluado. Aquí, de nuevo, aparecen dos mundos perfectamente diferenciados. Por un lado, los evaluadores centrados en la eficacia. Aunque la etapa tyleriana, como vimos, ya hace tiempo que se terminó, su influencia dura hasta nuestros días, convertida además en fuerza mayor de la mayoría de las evaluaciones. De ahí que éstas se hayan centrado sobre todo en la eficacia: el grado en que un acto propositivo logra lo que pretende. Por otro lado, los evaluadores centrados en la eficiencia. La meta de la eficiencia (que se propone descubrir el cómo se logra lo que se pretende -¿a qué precio?-), típica como veremos, entre otros, del modelo de *Análisis de sistemas*, pone a todas luces de manifiesto que éste tal vez sea el camino a seguir en el futuro para buena parte de los distintos enfoques evaluadores, aunque teniendo en cuenta tanto las formas duras (costes/beneficios) como sobre todo las más blandas (costes/efectividad), a la hora de evaluar los actos propositivos de carácter educativo.

En buena parte de las evaluaciones se realiza una *valoración de los efectos* de cada acto propositivo, aunque éste es un asunto de los más controvertidos de la evaluación formal, ya que existe todavía un grupo de evaluadores que piensan que la evaluación es equiparable a la investigación y, dentro de ésta en concreto, a la medición fiable y válida de los efectos, por lo que los juicios de valor debieran desaparecer de sus funciones. Esta postura, supuestamente aséptica y totalmente objetiva, no tiene hoy en día demasiados partidarios, ni siquiera entre los científicos de las ciencias duras (Stufflebeam, 2004). En el terreno de la evaluación se hace todavía más insostenible, pues al hablar de la necesaria evaluabilidad ya hice referencia a la necesidad de preguntarse por el tipo de objetivos propuestos, por el valor de los mismos. Si los objetivos son insustanciales no sería ético consumir recursos materiales y humanos en la evaluación formal, aun cuando ésta fuese técnicamente impecable (en el supuesto de que esto fuera posible para los humanos). Aquí residiría, por tanto, la clave para la distinción entre evaluación (emisión de juicios de valor) y medida. Por lo que se refiere a los efectos, se constata una enorme variación entre los proponentes de los distintos tipos de evaluación, ya que algunos de ellos sólo se preocupan de los previstos y propuestos, mientras que a otros, en el otro extremo, les interesa todo tipo de efectos, desde los establecidos o previstos a los imprevistos, sean unos y otros positivos o negativos.

Una última característica de la evaluación formal, especialmente relevante para el futuro, hace referencia a la propuesta de algunos evaluadores que consideran las actividades evaluadoras como un auténtico trabajo de investigación. Para estos autores, las

evaluaciones han de estar *relacionadas y guiadas por teorías*, de forma que aquéllas se conviertan en un instrumento básico de verificación o falsación de éstas. Sin duda, esto supondría un valor añadido importante a las cinco notas estructurales antes señaladas.

2.1. Tipos de evaluaciones

Cuando se habla de grandes tipos de evaluaciones, como los que pretendo destacar aquí, se constata que éstos vienen determinados por los tres momentos relevantes de cualquier acto propositivo: el inicio del mismo, su desarrollo y los resultados obtenidos tras su completa implementación. Esto es común a cualquier campo del saber, por lo que sus profesionales suelen seguir protocolos semejantes (véase Cuadro 1, parte central).

Gracias a la *evaluación inicial o valoración de necesidades* (también entrarían a formar parte de este gran tipo de evaluación las denominadas como evaluación diagnóstica, estudio descriptivo previo y análisis de la línea de base, aunque con diferencias manifiestas entre ellas en multitud de sentidos) es posible obtener una radiografía relativamente precisa de las necesidades existentes dentro de una determinada institución, de un individuo o de un grupo de personas determinadas. La aplicación de técnicas rigurosas nos proporciona una información relevante y bien fundamentada, gracias a la cual es posible la elaboración de una *jerarquía de necesidades*. Esta jerarquía resulta a todas luces imprescindible, pues lo ordinario es que existan muchas más necesidades que recursos materiales y humanos para satisfacerlas. Además, ocurre a veces que lo que piensan los máximos responsables académicos en los distintos niveles (ministerio, universidades, centros, institutos) sobre las necesidades a satisfacer no coincide con unas necesidades más realistas -sentidas-, sacadas a la luz gracias a la recogida de una información no sesgada (se tienen en cuenta las opiniones de todos los miembros que participan en los distintos contextos de aprendizaje/enseñanza: profesores, alumnos, personal de la administración y servicios y las propias autoridades). En la fase de la valoración de la evaluabilidad es donde se han de poner de manifiesto estas posibles discrepancias con el fin de tratar de darles una solución. Desde todos los puntos de vista parece lógico que si se han de iniciar acciones educativas concretas y sus pertinentes evaluaciones, aquéllas vengán condicionadas sobre todo por una jerarquía de necesidades avalada teórica y técnicamente, más que por modas o caprichos de quien en esos momentos pueda estar detentando el poder.

Realizado este primer paso imprescindible desde un punto de vista lógico, aunque la dura realidad nos muestre que más bien resulta infrecuente, el segundo es el de la *evaluación de procesos*. Al menos dos significados se suelen asignar al vocablo “procesos”. Por un lado, dentro de la psicología, cuando se habla de procesos se hace referencia a aquellos actos mentales (atención, percepción, memoria, pensamiento) que se supone son los responsables últimos de nuestros comportamientos. La evaluación necesariamente ha de ser indirecta, a través de ciertos indicadores de ejecución, pues no podemos realizar medidas directas de actos que sólo nos son accesibles a través de sus manifestaciones. Dentro del ámbito educativo, cabe afirmar que este tipo de evaluación es el que está hoy de moda cuando los objetos diana de la evaluación son los alumnos o los profesores. Es más, son bastantes los autores que ponen tal entusiasmo en los procesos que parece que los productos son irrelevantes. Consideran que éstos pertenecen a tiempos pasados y entienden que los procesos constituyen el auténtico termómetro de la modernidad y, por supuesto, de la postmodernidad. Mi planteamiento es que ambos son factores esenciales a tener en cuenta en los contextos de aprendizaje/enseñanza, pues la

única forma válida de poder evaluar los procesos es justamente a través de los productos de los cuales se supone son responsables. Mientras que al conductismo radical le bastaban los productos, pues no querían entrar en la caja negra de los procesos, creo que hoy estamos en condiciones de analizar también los procesos al contar con indicadores de ejecución, válidos y fiables. El otro significado de evaluación de procesos, posiblemente el más extendido entre los evaluadores, hace referencia a la valoración que se lleva a cabo a lo largo de la implementación de cualquier acto propositivo, en nuestro caso de tipo educativo. De esta forma, la evaluación nos proporciona una permanente retroalimentación de lo que va sucediendo momento a momento en el programa, a fin de que se pueda ir corrigiendo lo que no suceda según lo que estaba previsto. Es difícil poder dudar de la utilidad aparente de este tipo de evaluación de procesos, ya que ayuda a ir obteniendo una mejora continua de lo evaluado.

En los ámbitos educativos, al igual que en los ámbitos empresariales o clínicos, siempre partimos de unas metas u objetivos, dado que son ellos los que dan razón de ser de nuestras acciones. Parece lógico y consustancial al ser humano que éste quiera saber formalmente hasta qué punto se logró lo propuesto. Estaríamos ya hablando de *evaluación de productos*. Históricamente, no es de extrañar que este tipo de evaluación formal tuviese un predominio a lo largo de la primera mitad del siglo XX, gracias básicamente a la influencia tyleriana y conductista. Este planteamiento supuso sin duda un logro considerable con respecto a las evaluaciones informales típicas del siglo anterior, pues obligaba a los diseñadores de programas o proyectos a definir operativamente sus objetivos y a comprobar, al final, el grado de su ejecución. Con el predominio del enfoque cognitivista, en psicología, se puso tal énfasis en los procesos que casi se llegan a olvidar los productos.

Mi planteamiento es que se hace necesaria una llamada a la sensatez, avalada por los conocimientos científicos actuales, a fin de no intentar contraponer lo que es absolutamente complementario. No debiera olvidarse que lo que la sociedad demanda ante todo y sobre todo es la calidad de los productos (un buen profesor es el que ejecuta bien las tareas que tiene encomendadas; algo semejante vale para el alumno; ese tipo de alumno y profesor es el que quiere la sociedad). Ahora bien, si las variables independientes son los procesos, habrá, pues, que incidir en ellos por parte de los distintos agentes educativos (profesores, alumnos, padres, autoridades académicas, psicólogos educativos), aunque sin olvidar que lo que todos queremos al final es el mejor logro posible del producto que nos propusimos, aun cuando el producto final sea un proceso: razonar con corrección lógica o ser creativo, por ejemplo (Voss, Perkins y Segal, 1991). Estos procesos finales han de manifestarse y en su manifestación quedan automáticamente convertidos en productos. Resulta difícil que alguien pueda ser creativo si no se manifiesta en sus actuaciones (productos) creativamente. Este planteamiento de contemplar conjunta y complementariamente los procesos (en el doble sentido antes señalado) y los productos exige más esfuerzo y dedicación, pero a cambio permite obtener radiografías de lo evaluado mucho más fiables y válidas.

2.2. Principales enfoques

También aquí, al igual que ocurrió con las definiciones de evaluación, los plurales se imponen. Hoy por hoy no existe el enfoque óptimo, sino múltiples enfoques de evaluación, cada uno de ellos con sus ventajas y desventajas. Voy a descartar por razones obvias, las taxonomías de los enfoques considerados por Wolf (1969) como "blandos", aunque su utilización no sea en modo alguno tan infrecuente como a principios del siglo XXI cabría imaginar: el *modelo cosmético* (la apariencia del programa determina su bondad intrínseca); el *cardíaco* (la corazonada se constituye en criterio de valoración frente a los resultados); el *coloquial* (el equipo de "pensantes" determina la bondad de lo evaluado en un breve encuentro). También descartaré el otro extremo, el de las taxonomías que podríamos llamar "duras": las desarrolladas numéricamente (Williams, 1989), dada su escasa relevancia y visibilidad en los ámbitos profesionales. Me centraré en una de las taxonomías que goza de mayor reconocimiento internacional, la propuesta por House (1980), sabiendo que ello implica dejar de lado otras también bien conocidas por los evaluadores: algunas anteriores a la housiana (Popham, 1975; Stake, 1976a), otras algo posteriores (McCutcheon, 1982; Talmage, 1982; Stecher y Davis, 1987), aunque dentro de la misma década. Esta decisión hará que el lector pueda echar de menos, con toda razón, la exposición de ciertos enfoques que también han alcanzado reconocimiento internacional, como el modelo de evaluación democrática de MacDonald (1976), el de la evaluación iluminativa de Parlett y Hamilton (1977) o el de la evaluación sensible de Stake (1976b) y Guba y Lincoln (1982). Esta abundancia de taxonomías ya nos está indicando que, pese a no ser exhaustivas en la inclusión de todos y cada uno de los enfoques existentes, suelen ser de gran utilidad para quien se enfrenta por primera vez con la necesidad de tener que poner en práctica, con cierto rigor, un proceso determinado de evaluación. Lógicamente, se reduce de forma notable la relevancia de las mismas cuando nos adentramos en el mundo de los especialistas, pues no suele estar entre sus prioridades el añadir una más a la lista ya amplia de taxonomías de los enfoques existentes.

House (1980), tal cual aparece recogido en el cuadro adjunto, ha seleccionado los 8 modelos (que yo denominaré por razones de tipo científico y técnico enfoques) que cree han resultado más útiles dentro de la práctica profesional. A la izquierda del cuadro, en la primera columna, se muestran las denominaciones de los enfoques y en las 5 columnas siguientes, encabezadas por 5 parámetros (*Destinatarios, Consenso, Metodología, Resultados* y *Cuestiones Típicas*), las semejanzas y diferencias de estos enfoques. Creo que de estas cinco columnas, las tres últimas son las que nos pueden resultar más útiles a la hora de decidirnos bien por la elección de uno de estos enfoques o bien por uno nuevo, fruto de la selección armoniosa de los puntos fuertes de varios de ellos. La exposición que aparece a continuación de estos enfoques también puede verse en Fernández (2001).

Cuadro 2

Taxonomía (simplificada por mí) de los principales enfoques de evaluación (House, 1980)

MODELO	DESTINATARIOS	CONSENSO EN:	METODOLOGÍA	RESULTADO	CUESTIONES TÍPICAS
Análisis de Sistemas	Economistas	Metas, conocida relación causa-efecto	Variación planificada, análisis costes/beneficios	Eficiencia	¿Se han alcanzado los efectos previstos? ¿Se pueden conseguir más económicamente? ¿Cuáles son los programas más eficientes?
Objetivos Conductuales	Psicólogos	Objetivos preestablecidos	Tests de logro	Productividad	¿Logró el programa sus objetivos? ¿Rinde?
Toma de Decisiones	Administradores	Metas generales, criterios	Encuestas, cuestionarios, entrevistas	Control de calidad	¿Es el programa eficaz? ¿Qué partes son eficaces?
Libre de Metas	Consumidores	Consecuencias, criterios	Control de sesgos	Elección del consumidor	¿Cuáles son todos los efectos?
Crítica de Arte	Entendidos Consumidores	Normas	Revisión crítica	Normas perfeccionadas	¿Aprobaría un crítico este programa? ¿Ha aumentado la apreciación del público?
Revisión Profesional	Profesionales El público	Criterios	Revisión por jurado	Aceptación profesional	¿Cómo valorarían los profesionales este programa?
Cuasi-legal	Jurado	Procedimientos	Procedimientos cuasi-legales	Resolución	¿Cuáles son los argumentos en pro y en contra del programa?
Estudios de Casos	Clientes Profesionales	Negociaciones	Estudios de caso, entrevistas, observaciones	Comprensión de la diversidad	¿Qué les parece el programa a distintas personas?

El *enfoque de análisis de sistemas*, utilizado para evaluar actos propositivos complejos, con múltiples objetivos, algunos de ellos en claro conflicto entre sí, ha sido desarrollado principalmente, aunque no exclusivamente, por los economistas. Implica que las metas u objetivos han de estar claros, de forma que todos los posibles interesados puedan comprenderlos, desde los que se encuentran en los niveles más bajos hasta los máximos responsables institucionales, dado que difícilmente se puede trabajar en una misma dirección si hay ambigüedad en las metas. Desde un punto de vista científico y técnico, este enfoque presupone una relación causal entre variables independientes (causadoras) y dependientes (efectos). Dentro del marco educativo, tuvo su principal apogeo en la década de los 60 en Estados Unidos, con motivo de la implantación de la metodología de la *variación planificada* (enfoques y metodologías distintas a la hora de dar una solución a un determinado problema) en la familia de programas *Follow Through*. De lo que se trataba era de encontrar, dentro de un grupo amplio de propuestas cuyo objetivo era evitar el *desvanecimiento de los efectos* de los programas *Head Start*, aquéllas que ofrecieran una mejor relación productos obtenidos/recursos invertidos (humanos y materiales). La línea más dura, desde un punto de vista estrictamente metodológico, viene marcada por la utilización de los análisis de costes/beneficios, que implica la conversión, tanto de las variables independientes (causas) como de las variables dependientes (efectos), en términos monetarios. Los defensores de este enfoque hacen hincapié en la eficiencia (relación de entradas -recursos- y salidas -productos-), frente a la eficacia (sólo se tienen en cuenta los productos o efectos producidos: su grado de consecución). Por eso sus preguntas más características y, por tanto, más diferenciadoras con respecto a los otros enfoques son: ¿qué programa resulta más eficiente? o, dicho en otros términos, ¿cuál es el programa que ofreciéndonos los mismos resultados nos cuesta menos, tanto en términos de recursos humanos como materiales? Pese a este importante matiz diferenciador, es posible constatar que la influencia de la concepción tyleriana de la evaluación también se halla presente, como se refleja en una de las preguntas típicas: ¿cuál es el grado de consecución de los efectos previstos? Los puntos fuertes y débiles, en este enfoque como en el resto, están tan cercanos unos de otros que lo más fácil y sencillo es pasar de uno (el fuerte) a otro (el débil) sin darse apenas cuenta. El puente de transición, en todos los casos, suele ser la visión de túnel. Así, si bien puede resultar aconsejable que, dentro de una sociedad en la que casi todo se puede comprar y vender, se traten de transformar las variables independientes y dependientes, implícitas en el ámbito educativo, en términos monetarios, no es menos conveniente saber que no todo puede transformarse en puros términos económicos. Lo óptimo, pues, consistiría en, por una parte, ser consciente de lo que un sistema educativo cuesta y, por otra, saber que los costes son sólo un medio al servicio de unos determinados fines: la calidad o excelencia educativa y el desarrollo humano y social de los miembros que participan en todo sistema de aprendizaje y enseñanza. Una ampliación de las características teóricas y prácticas de este enfoque puede consultarse en Glasman y Biniaminov (1981), Rossi, Freeman y Wright (1979) o Tsang (1988).

El *enfoque de los objetivos conductuales* nos es relativamente bien conocido a los psicólogos, pues de la psicología procede, aunque ulteriormente haya sido asimilado y utilizado también por otros profesionales, tanto en el terreno educativo como empresarial. Se identifica plenamente con la concepción tyleriana de la evaluación: el hincapié recae, por un lado, en la concreción al máximo de las metas y, por el otro, en el logro de las mismas. De ahí que se requiera un claro consenso en los objetivos, que han de estar definidos desde un principio en términos operativos, y que se constituyen en la guía máxima de cualquier acto propositivo. Tanto desde el parámetro de la metodología como

desde el de los resultados siempre aparece la insistencia en la consecución de lo establecido (pruebas o tests de logro), en la productividad, es decir, en saber si se ha logrado lo que se pretendía. De aquí que sus preguntas típicas no puedan ser otras que las referidas al rendimiento, al logro, a la consecución de lo claramente previsto al inicio de cualquier tipo de acción emprendida. En cuanto a sus ventajas y desventajas, teniendo en cuenta lo ya referido en el enfoque anterior, cabe señalar que posiblemente su principal ventaja resida en los esfuerzos por aclarar al máximo los objetivos desde el comienzo y en los intentos de su valoración objetiva, aunque su desventaja más notoria tal vez se concrete en no tener en cuenta otros efectos posibles al no estar previstos, aun cuando pudieran ser más relevantes estos últimos que los primeros. Si se desea tanto una ampliación teórica como una implantación de este enfoque pueden verse Popham (1975) o Suchman (1967).

El *enfoque de la toma de decisiones* es el más anhelado por los administradores, los políticos, las autoridades académicas y, en general, por todos aquellos responsables sociales que tengan que tomar importantes decisiones que afecten a los demás, ya que debido a los datos que proporciona, al menos teóricamente, estarían más capacitados para llevarlas a cabo. Este enfoque implica un consenso en cuanto a metas y criterios, que se constituyen en los dos elementos básicos gracias a los cuales es posible hablar técnicamente del control de la calidad. Para el logro de este objetivo se van a utilizar técnicas muy variadas que son comunes a otros enfoques y a otras ramas del saber: encuestas, cuestionarios o entrevistas, entre otras. También en este enfoque la influencia tyleriana se hace sentir, como se manifiesta a través de la primera pregunta típica que hace referencia a la eficacia, es decir, a si se lograron o no los objetivos propuestos. Además, resulta interesante resaltar, por lo que respecta a la segunda pregunta, la más específica y diferenciadora -¿qué partes son eficaces?-, el hecho de que es en este enfoque donde primero aparece el desglose del acto propositivo en sus diversos componentes, a fin de poder diferenciar los que funcionan adecuadamente, según lo previsto, de aquellos otros cuyo rendimiento se aleja en mayor o menor grado de lo establecido (Stufflebeam, 2003). La posible visión de túnel puede aparecer cuando todo el hincapié se pone en la eficacia, no teniendo en cuenta los costes de los recursos humanos y materiales. Su posible mayor ventaja reside en el análisis minucioso de los distintos componentes de una determinada acción, con el fin de poder ir perfeccionando las partes que funcionan mal hasta acercarse al funcionamiento cuasi perfecto o alcanzar la denominada *calidad total* (Cornesky y McCool, 1992; Sánchez Matos, 2001).

El *enfoque libre de metas* es el que se sitúa en las antípodas de la visión tyleriana de evaluación. Como su denominación tan acertadamente nos sugiere, y nos ratifica su pregunta más significativa -¿cuáles son todos los efectos?-, se trata de realizar evaluaciones sin tener en cuenta prioritariamente los objetivos previstos por el autor de un acto propositivo. Es por ello por lo que puede resultar tan sumamente positivo para los consumidores, ya que debido a los resultados obtenidos, y siempre que se hayan llevado a cabo los oportunos controles de sesgos, ellos disponen de una visión muy precisa y exacta de aquello que ha sido evaluado. No es, por tanto, una casualidad que la mayoría de las asociaciones de consumidores de todo el mundo utilicen para sus evaluaciones de los más diversos productos que aparecen en el mercado, justamente este enfoque. Su ventaja más patente reside en la consideración de todos los efectos posibles -los previstos y los no previstos, bien sean unos y otros positivos o negativos-, que sin duda marca distancias considerables con respecto al resto de enfoques, la mayoría de ellos deudores de una u otra forma de la concepción establecida por Tyler. Su desventaja tal vez se concrete en una

cierta inviabilidad, en determinadas circunstancias, de esta característica que precisamente le diferencia de todos los demás enfoques, debido sobre todo a razones presupuestarias y de amplitud y altura de miras: desvelar todos los efectos. Uno de los autores más fervientes defensores de este planteamiento es Scriven (1974).

Ya dentro de una perspectiva más "cualitativa", House (1980) ha seleccionado también otros cuatro enfoques. El primero de ellos, el de la *crítica de arte*, recibe su denominación de los críticos de arte, por lo que cabe imaginar que las características esenciales de este enfoque se derivarán del modo de proceder de estos críticos (Eisner, 1979). Los dos pilares básicos sobre los que se asienta son, por una parte, el de los entendidos, tal cual se refleja en la primera pregunta típica -¿aprobaría un crítico/entendido este programa?- y, por otra, el de los consumidores, como se desprende de la segunda pregunta -¿ha aumentado la apreciación del público? Para que este enfoque funcione adecuadamente es imprescindible el consenso en los estándares o patrones, dado que de ellos depende la valoración final del acto propositivo. Una vez más las ventajas y desventajas aparecen muy juntas, pues si en su favor se puede esgrimir el juicio de un experto, en su contra puede aparecer la subjetividad incontrolada.

El *enfoque de la revisión profesional* es bien conocido para todo el que haya ejercido previamente una profesión, pues ha tenido que ser previamente acreditado para tal fin. En esencia se trata de confiar el valor de un acto propositivo al juicio, supuestamente bien formado, de un profesional especializado en esa materia. Se requiere un consenso en los criterios utilizados, dado que es sobre ellos sobre los que supuestamente se deberían basar las semejanzas de valor entre los profesionales. Dada la semejanza y, por tanto, el solapamiento entre este enfoque y el anterior, las ventajas (contar con alguien que por su experiencia sepa de qué está hablando) y desventajas (experiencia necesariamente limitada de cualquier especialista al campo concreto en el que hasta esos momentos ha estado trabajando) son también muy semejantes.

El *enfoque cuasi-legal*, recibe su inspiración y denominación del mundo de la judicatura (Wolf, 1974). Se trataría, en el terreno de la evaluación educativa, de imitar lo que sucede en un proceso judicial en el que nos encontramos con varios elementos esenciales: un juez que dicta sentencia, un acusado, un abogado defensor y un fiscal acusador. Ya dentro de nuestro campo de estudio, lo sometido a juicio sería un determinado acto propositivo, que contaría con defensores y detractores, ambos grupos exponiendo y defendiendo sus posturas contrapuestas, a fin de que en un último paso, una autoridad en la materia emitiera el juicio de valor definitivo, a la vista de los pros y los contras expuestos. Sin duda, con este enfoque es posible paliar al menos en parte algunas de las deficiencias que se pueden presentar en los dos anteriores (subjetividad incontrolada), pero uno de los problemas que puede presentar esta manifiesta ventaja es el de la duplicación de los recursos humanos y, por ende, de los costes.

El último enfoque, aunque no por ello el de menor valor, es el de *estudios de casos*. Su especificidad radica en la elección de un caso (proyecto, programa, curso, etcétera) que se presenta a muy diversas personas con muy distintas formaciones académicas y profesionales a fin de que dictaminen sobre el valor de lo que se les presenta a evaluación (Stake, 1978; Sutherland, 2004). Lo importante en este enfoque no es tanto elegir el mejor entre los diversos juicios, cuanto el asumir de forma integrada la diversidad. Volvemos, pues, a verificar el principio de que las ventajas y desventajas aparecen muy cercanas. Lo que se puede ganar en riqueza radiográfica de lo evaluado, al contar con muchas fuentes

de información, se puede perder en viabilidad: excesivo consumo de recursos materiales y humanos.

Aunque las taxonomías nos pueden ayudar a forjarnos una primera imagen certera de las especificidades de cada uno de los enfoques, académica y pedagógicamente hablando, la realidad poco a poco va imponiendo la necesidad de ir mezclando aspectos de cada uno de estos enfoques en subenfoques o supraenfoques, que gozan de delimitaciones y especificaciones mucho más precisas. Esto es lo que ocurre, por ejemplo, en el conocido "modelo de evaluación CIPP" -ver Cuadro adjunto-, elaborado originalmente por Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman y Provus (1971) y adaptado posteriormente por Payne (1982). La versión ofrecida aquí es una simplificación de la de Payne.

Cuadro 3

Modelo CIPP de Stufflebeam y colaboradores (1971), adaptado por Payne (1982) y simplificado por mí

	CONTEXTO METAS	INPUT DISEÑO	PROCESO ACTIVIDADES	PRODUCTO RESULTADOS
TOMA DE DECISIONES	¿Cuáles son las necesidades a satisfacer?	¿Qué diseño procesual se debería seguir para alcanzar los objetivos?	¿Se está implementando el diseño tal cual se había propuesto?	¿Resuelve el programa los problemas para los que fue diseñado?
RENDIR CUENTAS	¿Qué metas se eligieron cuando se inició el programa?	¿Por qué se eligió el diseño propuesto?	¿Qué pasos se dieron para ayudar a la gente a implementar el diseño del proyecto?	¿Cuáles fueron los efectos colaterales y los principales?

Este supraenfoque, y a la vez subenfoco –en función de las características que se deseen destacar-, se enmarca dentro de la perspectiva general de la *toma de decisiones*. Sus tres notas más específicas y distintivas son:

- el hincapié que se hace en que los datos de la evaluación deben servir ante todo y sobre todo para tomar decisiones bien fundamentadas;
- el hecho de que la evaluación es uno de los mecanismos esenciales a la hora de tener que rendir cuentas ante quien ha financiado el acto propositivo;
- la necesidad de considerar conjuntamente los cuatro componentes esenciales de este tipo de evaluación: el contexto o las metas, el input o el diseño, el proceso o las actividades y el producto o los resultados (Stufflebeam, 2003).

Para cada casilla se ha seleccionado una única pregunta de las múltiples que aparecen en el modelo adaptado de Payne, a fin de proporcionar indicios válidos de la estructura completa de todo este submodelo. Este tipo de submodelos o supramodelos eclécticos o integradores tiene su razón de ser en que la realidad siempre suele desbordar las visiones subyacentes a cada enfoque, por lo que los profesionales evaluadores sienten la necesidad de tener en cuenta conjuntamente distintas características que aparecen clasificadas en distintos enfoques y que son consideradas esenciales para su práctica. En este sentido, cabría indicar que si bien las taxonomías son condición necesaria para un primer acercamiento al campo de la evaluación, en modo alguno se constituyen en condición suficiente para la práctica cotidiana del evaluador. En la evaluación real, las circunstancias concretas pueden requerir que se unan ciertas características que en el nivel teórico aparecen clasificadas en lugares distintos.

3. Las metaevaluaciones

Una vez realizada la evaluación, inspirada en uno u otro modelo o en submodelos/supramodelos, ha de hacer su aparición la metaevaluación (Scriven, 1969) que, una vez más por desgracia, no ha sido suficientemente valorada por los propios evaluadores y, por consiguiente, apenas se ha puesto en funcionamiento, pese a la evidente importancia que tiene dentro de todo el proceso evaluador (Straw y Cook, 1990). El objetivo fundamental de la metaevaluación, como ya su propio nombre indica, es el de evaluar las evaluaciones: su calidad técnica y lo acertado de sus conclusiones (Stufflebeam, 2001). Esto implica que las denominadas *evaluaciones primarias* van a ser examinadas para detectar sus posibles sesgos, sus errores técnicos, sus dificultades administrativas o sus usos indebidos, como acertadamente señala Stufflebeam (1974). Es preciso recordar aquí que una de las principales implicaciones de la grave crisis por la que pasó la evaluación en la etapa del realismo o de la expansión fue que ésta no era todo lo rigurosa que en principio cabía esperar, por lo que las conclusiones que se derivaban de ella podían producir efectos contraproducentes.

Una de las vías más eficaces para paliar estos efectos no deseados es, precisamente, la realización de metaevaluaciones (Ory, 1992). Una síntesis de los principales criterios a utilizar dentro del proceso metaevaluador es la ofrecida por Payne (1982), que enmarca

dichos criterios dentro de cuatro grandes bloques, tal cual queda reflejado en el Cuadro adjunto.

Cuadro 4
Criterios de la metaevaluación

DISEÑO GENERAL DE LA EVALUACIÓN	RECOGIDA Y PROCESAMIENTO DE LOS DATOS	COMUNICACIÓN, PRESENTACIÓN E INFORME	PRUDENCIA Y PROPIEDAD
ALCANCE	REPLICABILIDAD	OPORTUNIDAD	ETICIDAD
RELEVANCIA	OBJETIVIDAD	TOTALIDAD	PROTOCOLO
FLEXIBILIDAD	REPRESENTATIVIDAD		
VIABILIDAD			

Lo primero que preocupa a quien realiza una metaevaluación es todo lo referido al diseño de la evaluación (evaluación primaria) que va a ser objeto de valoración. Con el fin de poder emitir un juicio certero sobre el mismo, el metaevaluador ha de apoyarse en cuatro criterios fundamentales:

- el del *alcance*: ¿se han tenido en cuenta todos los aspectos significativos en el plan de evaluación y las variables que pudieran afectar negativamente a la evaluación?;
- el de la *relevancia*: ¿son todos los datos recogidos relevantes para los afectados y para los objetivos propuestos?;
- el de la *flexibilidad*: ¿goza de suficiente flexibilidad como para tener en cuenta los cambios ocurridos en el programa desde que fue diseñado hasta que se acabó su implementación?;
- el de la *viabilidad*: ¿se elaboró un diseño realista, prudente y ajustado, en cuanto al presupuesto?

Analizado el diseño, el paso siguiente es el examen detenido del tipo de recogida y procesamiento de los datos que se ha llevado a cabo en la evaluación. Para ello se recomienda utilizar un trío de criterios:

- el de la *replicabilidad*, centrado en el control de instrumentos y procedimientos a fin de asegurar su posible réplica en ulteriores estudios;
- el de la *objetividad*, dirigido a la detección y el control de los sesgos;
- el de la *representatividad*, referido a procedimientos y métodos.

A continuación, el metaevaluador ha de detenerse lógicamente en el apartado de la presentación de los resultados de la evaluación, mediante los correspondientes informes. Dos son aquí los criterios a tener en cuenta:

- el de la *oportunidad*, es decir, constatar si la comunicación de los datos se hizo en el momento oportuno para que se pudieran tomar las pertinentes decisiones;
- el de la *totalidad*, es decir, asegurarse de que se entregaron tantos informes cuantos eran necesarios, en función de los diferentes destinatarios o potenciales beneficiarios de los datos.

En último lugar, hay que detenerse en la valoración de las consideraciones éticas y legales que han de estar presentes en cualquier proceso evaluador. Los dos criterios a utilizar aquí por el metaevaluador son:

- el del *cumplimiento de la normativa ética* relacionada con los derechos de privacidad y confidencialidad;
- el del *protocolo*, que implica el seguimiento escrupuloso de los pasos marcados socialmente para los estudios de evaluación, de forma que no se destruya la cooperación futura.

De forma mucho más abarcadora, Stufflebeam (1978) nos ofrece una visión general sobre la metaevaluación, vertebrada en torno a dos ejes:

- el de los *patrones*: principios consensuados por evaluadores profesionales para medir el mérito de una evaluación;
- el de las *pautas*: indicaciones de procedimiento que ayuden a los evaluadores a satisfacer los patrones de evaluación.

Este autor da cuenta de 34 patrones y de 68 pautas, que presenta a la comunidad científica para su debate y ulterior ratificación o rectificación. Igualmente, es posible encontrar pistas abundantes sobre metaevaluación en los volúmenes del *Joint Committee on Standards for Educational Evaluation* (1981, 1988, 1994, 2003) o en el trabajo del *ERS Standards Committee* (1983), aunque en cualquiera de estos tres últimos casos es bastante probable que el lector pueda sentirse algo perdido, por lo que, si desea que le den el trabajo ya hecho, puede detenerse en la síntesis sobre el papel del metaevaluador y de los usos formativos y sumativos de la metaevaluación que han elaborado Worthen y Sanders (1987).

Estos tres organizadores conceptuales -valoración de la evaluabilidad, evaluación (inicial, de procesos y de productos) y metaevaluación-, que hasta el presente han sido utilizados la mayoría de las veces de forma independiente y descoordinada, aparecen hoy en día más bien como partes complementarias de un mismo complejo proceso evaluador. Es más, cada vez va pareciendo más pertinente que, dentro del bloque de la evaluación, se incluya también otro componente, complementario de los otros tres, denominado *evaluación de seguimiento* (valorar la eficacia/eficiencia de un acto propositivo en tiempos

posteriores a la primera evaluación, que suele llevarse a cabo una vez éste ha concluido). Gracias únicamente a este tipo de evaluación es posible detectar si los efectos, manifiestos en un determinado momento –primera evaluación–, se van desvaneciendo a medida que pasa el tiempo, permanecen estables o, como en ocasiones sucede, se van acrecentando. Es obvio que el juicio de valor de referido al acto propositivo puede cambiar considerablemente, en función precisamente de los datos de esta evaluación de seguimiento, con respecto al forjado tras la primera evaluación.

4. Comentarios

Con harta frecuencia se olvida que el proceso evaluador se puede representar como un continuo que es considerablemente más extenso que la ya de por sí compleja evaluación. De hecho, en uno de sus extremos ha de aparecer la valoración de la evaluabilidad. Gracias a ella es posible ahorrarnos sorpresas desagradables: intentar valorar lo no evaluable, servir intereses encubiertos no deseados u ocasionar consumos excesivos de recursos materiales y humanos, entre otros. En su faceta más positiva, la valoración de la evaluabilidad nos permitirá preparar adecuadamente el terreno para que pueda germinar adecuadamente la evaluación en cuanto tal.

En el centro del continuo se producirán los distintos tipos posibles de evaluación: inicial o de necesidades, de procesos y de productos, junto con la evaluación de seguimiento. Es importante señalar aquí, que desde la perspectiva adoptada en este libro, no hay cabida para la tan de moda contraposición entre productos y procesos. Ambos se requieren mutuamente, a la par que se completan con la evaluación inicial. Gracias a ésta podemos percatarnos de las características del terreno sobre el que se edificará nuestro acto propositivo. Mediante la evaluación de procesos podemos ir sabiendo hasta qué punto se da una concordancia entre lo diseñado y lo implementado. Debido a la evaluación de productos podemos conocer, con cierto nivel de certeza, el grado de consecución de nuestros objetivos o metas. El conocimiento del desvanecimiento, la permanencia o el incremento de los efectos nos vendrá dado por la evaluación de seguimiento.

Por fin, resulta imprescindible la reflexión sobre la propia evaluación, es decir, la metaevaluación, dada la importancia puesta en la información recogida por la evaluación para las decisiones a tomar. Si los datos son sesgados, si la interpretación es parcial, si el diseño no es el correcto, la información obtenida mediante una evaluación primaria nos llevará a tomar decisiones tal vez peores que las que se hubiesen podido realizar sin contar con los resultados de ésta. Gracias a la metaevaluación nos es dado saber si la evaluación ha resultado ser más bien cosmética o puramente burocrática o, por el contrario, se ha enmarcado dentro de una cultura sentida de evaluación, por lo que los distintos agentes del ámbito educativo pueden comenzar a poner en práctica las implicaciones derivadas del proceso evaluador.

A la luz de lo indicado sobre la valoración de la evaluabilidad, los distintos tipos de evaluación y la metaevaluación, cabe afirmar que en cualquier proceso de evaluación resulta imprescindible la presencia de personas especializadas (profesionales de la evaluación) para realizar con éxito tan complejo trabajo en cualquiera de los niveles educativos: educación infantil, primaria, secundaria o universidad.

CAPÍTULO TERCERO

Las evaluaciones centradas en el profesorado a lo largo de su carrera profesional

1. Introducción

La evaluación del profesorado, tanto en los niveles no universitarios como universitarios, ha de enmarcarse dentro de la cultura de la evaluación a la que se hizo referencia en el primer capítulo y muy concretamente en el contexto de la profesionalización del profesorado y la reestructuración de las instituciones. A su vez, las evaluaciones han de entenderse teniendo en cuenta lo señalado en el segundo capítulo sobre las valoraciones de la evaluabilidad, las evaluaciones iniciales o de necesidades, de procesos, de productos, de seguimiento y, en último lugar, las metaevaluaciones. El manual de referencia obligada sobre evaluación del profesorado es sin duda el de Millman (1981), así como el que aparecerá nueve años más tarde, en esta ocasión sólo para los niveles no universitarios: educación primaria y secundaria (Millman y Darling-Hannmond, 1990). En el Prefacio, estos autores ya se encargan de indicarnos el cambio considerable producido a lo largo de esa década de los 80 en la evaluación del profesorado, por lo que el segundo manual, nos indican, será muy diferente del primero, tanto en autores como en contenidos. La razón dada es la acomodación a los cambios producidos en ese corto espacio de tiempo. Estos cambios se han reflejado en la aparición de nuevas revistas especializadas (*Journal of Personnel Evaluation in Education*), en los monográficos aparecidos o en el considerable volumen de artículos publicados, por sólo citar algunos ejemplos ilustrativos. Desde un punto de vista conceptual el cambio más radical viene dado por la conciencia de la importancia de la evaluación del profesorado para su mejora y la de la institución docente. Dado que en el pasado relativamente reciente no había aflorado tal conciencia –de hecho algunos autores nos hablan de la *asignatura pendiente* en educación-, no cabía imaginar unas actividades evaluadoras del profesorado en los ambientes educativos. Es más, la certificación u oposición de entrada, de carácter vitalicio, difícilmente se podía compaginar con evaluaciones ulteriores. ¿Qué sentido podían tener éstas? ¿No sería como una especie de oxímoron? Paulatinamente, esta mentalidad de tipo funcionarial –el examen acreditativo de entrada concede validez para el ejercicio docente hasta la jubilación- va a ir siendo sustituida, al menos en los Estados Unidos, por otra en la que la temporalidad cobra pleno protagonismo, de forma que la certificación inicial se ha de ir poniendo a prueba mediante las correspondientes evaluaciones posteriores.

Dentro de esta nueva mentalidad –*provisionalidad de acreditaciones y certificaciones iniciales*-, las evaluaciones del profesorado están llamadas a cumplir varias funciones, que van desde la selección de entrada, hasta la posibilidad de rescisión del contrato por incompetencia manifiesta, pasando por la utilizada para su propia formación y mejora en su ejecución dentro de las aulas. Ésta es, hoy por hoy, la función más destacada y la más aceptada tanto por los evaluadores como por los evaluados. A su vez, la experiencia acumulada a lo largo de más de un siglo nos indica que estas evaluaciones pueden ser muy variadas metodológicamente hablando (recuérdese lo visto en el capítulo segundo), con el fin de que se ajusten al máximo a las características peculiares del contexto en donde se llevan a cabo. No es necesario ser un especialista consumado en evaluación para saber que los tipos de evaluación no pueden ser los mismos dentro del ámbito universitario que en educación infantil o primaria. Pero, incluso dentro de los mismos niveles –esto es lo que nos ha ido enseñando la experiencia-, los mismos tipos de evaluación pueden producir resultados contrapuestos, por lo que se hace necesaria una plena acomodación del tipo de evaluación a las demandas específicas de cada contexto.

2. Los fines de la evaluación

Muy probablemente, una de las primeras preguntas que surgen cuando se quiere implantar cualquier sistema de evaluación sea la de los objetivos de la misma. Todo lo que realicen, tanto los evaluadores (diseño, instrumentos de valoración, tipo de informe) como los evaluados (actitud positiva o receptiva frente a pasividad o incluso rechazo), vendrá condicionado por la respuesta dada a esta crucial pregunta. Natriello (1990) nos señala tres propósitos de la evaluación. El primero hace referencia al *control del rendimiento* para lograr una enseñanza eficaz, reforzando los comportamientos individuales de alto rendimiento con incrementos salariales o diversos tipos de pluses, si fuera necesario. La lógica de este planteamiento es bien conocida en nuestros días, sobre todo por la influencia de los patrones vigentes en el mundo empresarial. Se trataría de que cada profesor, en su puesto, desempeñe unas determinadas funciones en su grado máximo de excelencia, para lo cual necesita una retroalimentación que le informe de los posibles puntos débiles y le ofrezca pistas para superarlos. Ésta sería la labor encomendada a la evaluación: centrarse en la mejora del rendimiento de cada profesor, sin realizar modificaciones del sistema. Los efectos directos esperados o incluso los no esperados (estrés o ansiedad) se circunscriben al ámbito del individuo (al profesor). Aunque, sin duda, también se pueden producir efectos indirectos deseados o indeseados desde un punto de vista institucional.

El segundo propósito hace referencia a la *oportuna movilidad dentro del sistema*. Gracias a la evaluación, se podría saber quiénes son los individuos deseosos de alcanzar una determinada posición dentro del sistema –un ascenso-, los que piensan abandonarlo o aquellos a los que sería conveniente exigirles su retirada. Aquí el propósito es la introducción de cambios en el sistema mediante la movilización de sus profesores. Los efectos previstos van, pues, dirigidos a un cambio organizativo, con la intención de mejorar el rendimiento global, dejando de lado la preocupación por el rendimiento concreto de cada profesor.

El tercer propósito se refiere al *control del sistema organizativo* en cuanto tal. Mediante las evaluaciones se pretende dar a conocer al profesorado y a la sociedad en general que las decisiones tomadas por la jerarquía institucional son justas, racionales y equitativas, pues están fundamentadas en informaciones recogidas de forma fiable y válida. Los efectos pretendidos van más allá del propio centro escolar, pues con la evaluación se pretende rendir cuentas ante aquellas autoridades académicas y sociales que han otorgado el dinero de los contribuyentes a unos centros para que éstos logren unos fines educativos específicos. Lo dicho valdría igual para los centros públicos que para los privados, es decir, para aquéllos financiados únicamente con fondos aportados por entidades privadas.

3. Aspectos a evaluar

En principio, parecería razonable evaluar todos aquellos aspectos que van a tener una incidencia en la calidad del proceso de aprendizaje/enseñanza. No cabe duda de que el nivel de conocimientos es una condición necesaria aunque no suficiente para poder ser un buen profesor. Ahora bien, esta dimensión hay que enmarcarla en cada era o etapa histórica. Hasta hace poco esta característica resultaba tan relevante que durante un tiempo se llegó a creer que era la única que merecía la pena tenerse en cuenta, a la hora de evaluar a un profesor para su contratación. Hoy, en la sociedad del

conocimiento y la información, esta faceta ha de ser necesariamente integrada dentro de un conjunto de competencias que han mostrado ser relevantes a la hora de explicar la calidad del producto/proceso educativo. Por eso en nuestros días, además de los conocimientos e incluso por encima de ellos, se hablará de las capacidades en tanto *habilidades*, gracias a las cuales el profesor es capaz de inferir, deducir, clasificar, sintetizar (dentro del ámbito cognoscitivo), orientarse perfectamente en lo espacial y temporal (dentro del terreno de lo psicométrico), expresarse adecuadamente (dentro del área de la comunicación), o participar e integrarse con el resto de agentes docentes y discentes (dentro del campo de la inserción social). Estas capacidades pueden desglosarse en competencias (*repertorio de capacidades*), en tanto hacen referencia al grado de dominio que muestra un profesor. Junto con la consideración de las dimensiones más cognoscitivas hoy en día se hace igualmente hincapié en el mundo de los valores, las creencias y las actitudes. Es difícil concebir la existencia de un buen profesor, en el sentido más completo e integrado del término, con escasa valoración del desarrollo de las personas –sus alumnos–, con una actitud negativa hacia el conocimiento o con unas creencias sexistas, racistas o xenófobas.

Hasta aquí, la evaluación se circunscribe a las variables intraorganísmicas, pues tanto los conocimientos como las capacidades, competencias, valores, creencias y actitudes sólo son concebibles dentro de una persona. A estas variables habría que añadir, ahora, el fruto de la puesta en funcionamiento de precisamente estas variables independientes. Me estoy refiriendo al rendimiento (del alumno, del centro), principal variable dependiente, que viene condicionada y determinada por las variables intrínsecas del profesor, por las variables específicas del alumnado y por los condicionantes contextuales, como se puso de manifiesto en el capítulo primero (Figura 1). Estrechamente relacionada con el rendimiento aparece la eficacia (logro de los objetivos propuestos) y la eficiencia (relación entre logros y recursos materiales y humanos puestos en funcionamiento) del proceso de aprendizaje/enseñanza.

4. Evaluación del profesorado en sus distintas etapas profesionales

En todos los países la carrera profesional del profesorado pasa por diversas *etapas*: certificación o acreditación necesaria para el acceso –un determinado título–, programas de formación o capacitación, selección del profesorado, profesorado contratado, profesorado funcionario o fijo y dentro de esta categoría, ascenso a distintos niveles. Cada una de estas etapas, en principio, requiere algún tipo de evaluación, pues sobre los resultados de esta evaluación es sobre los que se cimentarán las correspondientes decisiones que denieguen o den paso a cada una de las siguientes etapas.

Comenzaré por los *títulos*. En general, se puede afirmar con cierto rigor que las calificaciones universitarias por sí mismas no permiten predecir el grado de eficacia del futuro profesorado. El fundamento para tal aserto se basa tanto en las investigaciones realizadas al respecto (Pratt, 1987), como en los datos provenientes de los que aspiran a llegar a ser un día profesores. Mi experiencia como evaluador de licenciados aspirantes a profesores de educación secundaria en España, durante varios años, indica que si bien los contenidos aprendidos durante la carrera son juzgados suficientes para su futura enseñanza, no ocurre lo mismo con respecto al cómo impartirlos. El aprobado, notable o sobresaliente en contenidos se convierte así en *condición necesaria*, pero en modo alguno en *condición suficiente*.

Si continuamos con los *programas de formación o capacitación*, el panorama que se nos presenta allá donde dirijamos la mirada no parece muy halagüeño. Según el informe Feistritz (1984), más de la mitad de los 1.287 programas de formación de profesorado en los Estados Unidos deberían ser clausurados. En España, los datos con los que contamos irían en la misma dirección, cuando se analizan los resultados del Curso de Aptitud Pedagógica (CAP), que es todavía hoy obligatorio para todos aquellos que aspiran a ejercer de profesores de secundaria (ver Tabla 1).

Tabla 1
Resultados sobre la necesidad y utilidad de los cursos de formación o capacitación del profesorado (CAP) en España

Años	Opciones	Ítems				
		Supresión del CAP	Útil para formar profesores	Necesidad preparación distinta CAP	La licenciatura es suficiente	CAP realizado por ser obligatorio
	Sí	202 (78.6%)	34 (13.2%)	171 (66.5%)	97 (37.7%)	208 (80.9%)
1982*	No	25 (9.7%)	192 (74.7%)	39 (15.2%)	129 (50.2%)	23 (9%)
	Sin resp.	30 (11.7%)	31 (12.1%)	47 (18.3%)	31 (12.1%)	26 (10.1%)
	Sí	140 (77.8%)	10 (5.5%)	150 (83.3%)	80 (44.4%)	176 (97.8%)
1983**	No	10 (5.5%)	170 (94.5%)	30 (16.7%)	100 (55.6%)	4 (2.2%)
	Sin resp.	30 (16.7%)				
	Sí	91 (91%)	19 (19%)	95 (95%)	34 (34%)	90 (90%)
1984***	No	9 (9.5%)	81 (81%)	5 (5%)	66 (66%)	10 (10%)

* (N=257); ** (N=180 elegidos al azar de un total de 426); *** (N=100 elegidos al azar de un total de 432). Todos los participantes eran licenciados, en distintas Titulaciones. El CAP se impartió en el Instituto de Ciencias de la Educación de la Universidad Complutense de Madrid.

Aquí nos encontramos con una aparente paradoja, pues mientras la realidad de lo que se hace y da en estos cursos parece ser enjuiciada bastante negativamente, los aspirantes a profesores, junto con los investigadores educativos, reconocen que necesitan una determinada y específica formación, además de la proporcionada en la carrera, para poder enfrentarse con éxito a sus futuras funciones como profesor (Good y Weaver, 2003; Hoy y Spero, 2005).

Si contamos con estudiantes con su carrera terminada y que ya han pasado los programas de formación o capacitación, pues desean ser profesores, el paso siguiente es el de la *selección*. Como señala Scriven (1990), la tarea de selección consiste en:

- elegir al mejor candidato -tarea de clasificación-;
- decidir si un candidato es suficientemente bueno como para que se le asigne un determinado puesto -tarea de calificación-;
- una combinación de ambas.

En cualquiera de los tres casos una de las cuestiones básicas es la referida a *qué deberíamos buscar*, cuya respuesta exige que se tengan en cuenta la misión y necesidades de la institución, junto con las funciones a desempeñar y una descripción minuciosa del trabajo concreto que se habría de desarrollar. Los conceptos clave a utilizar, según este autor, serían el de *mérito*, valorado según los pertinentes estándares profesionales (productividad y rendimiento pasados), el de *valor* que tiene en cuenta las necesidades completas de la institución para la que se va a trabajar y el de *utilidad*, derivado de la consideración ponderada de los dos conceptos anteriores. Se requiere un equilibrio variable de valor y mérito, en función de las especificidades de cada contexto. En general, las cualidades que se deberían buscar, según Scriven, son: unos conocimientos sólidos en el área del saber que se ha de impartir, competencia en evaluaciones curriculares y del alumnado, capacidad para enseñar en clase, cualidades intelectuales y personales relacionadas con la enseñanza y otros valores de utilidad para el centro o la institución.

A partir de aquí surgiría la siguiente cuestión: *cómo buscar*, es decir, con qué tipo de técnicas. Casi todos los especialistas en evaluación juzgan oportuno la *utilización conjunta de varias técnicas de recogida de la información* (análisis de los currículos, informes o notas de los programas de formación, entrevistas, resultados de tests, observaciones, etcétera). Obtenida la pertinente información, la tercera cuestión se refiere a lo que deberíamos hacer con la información, es decir, cómo la combinamos en aras de reducir la probabilidad de errores en las decisiones a tomar. Esto requiere trabajar con múltiples variables ponderadas según una jerarquía que ordena esas variables en función de su supuesta relevancia a la hora de predecir el futuro rendimiento del profesor. Por fin, la cuarta cuestión se refiere a quién debería hacer todo lo anterior. Parece lógico que sea un equipo entre cuyos miembros estén profesores en ejercicio, autoridades académicas y especialistas en evaluación.

Una vez seleccionado, el profesorado debiera ser evaluado en sus *primeros años de ejercicio profesional*, con el fin de:

- retener a los profesores más prometedores, en el supuesto de que la selección no haya tenido un carácter permanente;
- ayudarles, en todo caso, en sus iniciativas, en sus aportaciones y en su propia autorregulación (Peterson, 1990).

En los primeros años estos profesores necesitan mucha ayuda para su crecimiento profesional, para su correcta socialización y, sin embargo, la

retroalimentación de tipo evaluativo sobre su trabajo suele ser en la realidad muy escasa. Con el fin de paliar estas manifiestas deficiencias, este autor propone una evaluación basada en los siguientes pilares. El primero de ellos hace referencia a las *fuentes múltiples de datos* (tests para profesores, informes de la dirección, observación sistemática, revisiones de los materiales empleados por parte de los colegas, entre otras). Aunque, en principio, el coste pueda parecer elevado, los beneficios a largo plazo parecen considerables. El segundo se refiere a la preparación de un *dossier* en el que se ponderen específicamente los diferentes datos procedentes de las distintas fuentes, de forma que se pueda obtener un adecuado perfil sobre los puntos fuertes y débiles de cada profesor. El tercero requiere la *implicación del profesor en su propia evaluación*, pues la unión de la perspectiva de los demás y de su propia perspectiva ayuda a tener una visión más completa y correcta de lo que sucede con un determinado profesor en un contexto concreto de aprendizaje y enseñanza. El cuarto implica la participación de *múltiples juicios* sobre el rendimiento del profesor, con el fin de evitar al máximo los sesgos provenientes de una única fuente, como puede suceder cuando la evaluación se confía, por ejemplo, al director del centro únicamente.

Admitido e integrado el profesor en cada institución o centro docente, el paso siguiente a dar sería el de tratar de ir más allá de la competencia mínima (adquirida en los primeros años de ejercicio) mediante el *desarrollo profesional*, entendido éste como el proceso por el cual cada profesor va alcanzando niveles más altos de competencia: desarrollo pedagógico, profesional, organizativo o personal. Uno de los medios esenciales para el logro de este desarrollo es sin duda la evaluación. Ahora bien, no cualquier tipo de evaluación por sí misma va a contribuir a este desarrollo profesional. Según Duke y Stiggins (1990), para que esto ocurra es necesario tener en consideración tres tipos de características. Por un lado, las de los *profesores* (fuertes expectativas profesionales, orientación positiva hacia los riesgos, deseo de experimentar en clase, alguna experiencia anterior positiva con la evaluación, etcétera), junto con las de los *evaluadores* (facilitadores de crecimiento: credibilidad, actitud cooperativa, confianza, base de conocimientos para la mejora, etcétera). Por otro, las características del *proceso evaluador*: claridad en todos y cada uno de los pasos a lo largo del proceso, conocimiento compartido de evaluadores y evaluados, adecuación de las observaciones realizadas, retroalimentación útil y significativa. En tercer lugar, las características del *contexto*: evaluación dirigida al crecimiento de la competencia profesional, dedicando todo el tiempo que sea necesario, tanto por parte del evaluador como del evaluado, al desarrollo de estas competencias. Los responsables institucionales (los directores de los centros sobre todo) han de cumplir un papel muy relevante a la hora de estimular y potenciar un ambiente propicio para este crecimiento.

Además de contribuir al desarrollo profesional de los profesores en ejercicio, la evaluación ha sido considerada un instrumento útil para las *promociones, subidas salariales o incluso para los casos de cese de la contratación*. De esta forma la evaluación muestra su otra finalidad: *la denominada sumativa*. Mientras que no suele ocasionar reticencias insalvables cuando la evaluación se utiliza con su finalidad formativa (mejora de lo evaluado, en nuestro caso del profesor en sus diversas etapas profesionales), no es posible decir lo mismo con esta otra finalidad. Cierta tipo de profesorado, con o sin razón, intentará poner mil y una pegas a cualquier proceso evaluador de este tipo. Esta finalidad sumativa se materializará en dos versiones bien conocidas: el *salario según méritos*, en función de la evaluación de su rendimiento y las *escalas profesionales*, según una jerarquía previamente establecida que marca los

escalones por los que se puede ir ascendiendo a partir de determinados méritos acumulados. Aquí la evaluación requiere el establecimiento de unos estándares a partir de los cuales se van a tomar las decisiones. Los mínimos nos permitirán diferenciar entre profesorado competente e incompetente, mientras que los estándares competitivos posibilitan tanto las discriminaciones salariales como los ascensos profesionales. No son pocos los autores que se preguntan si sería posible armonizar este aspecto sumativo de la evaluación con el formativo, pues de esta forma muchos de los recelos y oposiciones de aquél tal vez desaparecerían. Bacharach, Conley y Shedd (1990), de hecho, pretenden transformar los sistemas de salarios según méritos y escalas profesionales en planes de desarrollo profesional. Según estos autores, dicha transformación únicamente se puede dar dentro de un contexto en el que se reduzcan al mínimo la *cultura burocrática y la atmósfera conflictiva* que tan frecuentemente caracterizan a los centros actuales. En estas circunstancias la desconfianza hace acto de presencia, imposibilitando de raíz la cultura de la evaluación para el desarrollo profesional.

Hasta ahora, la evaluación, bien de tipo formativo o sumativo, o bien armonizadas ambas funciones, ha estado centrada en los individuos: en cada uno de los profesores. Un paso más es el que nos encamina hacia una evaluación centrada en la *calidad o excelencia de la institución en cuanto tal: la mejora de la escuela, los colegios, los institutos o las universidades*. Desde este planteamiento, cobra especial significación la evaluación que se detiene en el fenómeno hasta ahora relegado a segundo plano: *el de las relaciones entre los distintos agentes de cualquier sistema de aprendizaje/enseñanza (las que mantienen los profesores entre sí, las de éstos con sus alumnos; las de los profesores con las autoridades académicas), sabiendo que todas estas relaciones siempre han de ser consideradas desde la bidireccionalidad*. Vistas así las cosas, este tipo de evaluación resulta complementario del resto de evaluaciones centradas en los individuos. Cuando el hincapié recae en la institución, el evaluador ha de planificar sus evaluaciones teniendo en cuenta las necesidades de los alumnos, las motivaciones de los profesores y los productos globales surgidos de las interacciones entre ambos colectivos. El desarrollo de estos dos tipos de agentes (en conocimientos, capacidades, valores, actitudes) ha de verse como parte esencial del desarrollo de la propia institución educativa (recuérdese lo dicho en el capítulo primero: Figura 1). Se trata de unir esfuerzos en una misma dirección, pues el desarrollo (la mejora) necesariamente es compartido. Se da un amplio solapamiento de mejoras individuales que conducen a la calidad o excelencia total o institucional.

5. Distintos procedimientos de evaluación

Si como hemos ido viendo, hay varios enfoques de evaluación y además se pueden aplicar con finalidades diversas, cabe esperar que sean muchos los procedimientos concretos de evaluación con los que se pueda contar para realizar la evaluación del profesorado. Aquí dejaré de lado aquellos centrados en lo que es la *certificación* (los requisitos exigidos para obtener un grado o licenciatura en donde se supone se han obtenido los conocimientos necesarios para ejercer la profesión de profesor), o la *acreditación* (superación de unos estándares más o menos rigurosos elaborados por las correspondientes instituciones -agencias de profesionales o ministeriales, centros de evaluación-), al estar ya suficientemente reglamentados en los distintos países (Berk, 1986). Esto no significa, en modo alguno, que se den sin más por buenos y válidos tales procedimientos. Es más, el corpus bibliográfico al respecto parece indicarnos que buena parte de los existentes son a todas luces insuficientes y,

además, con considerables deficiencias, cuando se utilizan como únicos mecanismos para garantizar la calidad del profesorado (Millman y Darling-Hammond, 1990).

Ninguno de los procedimientos utilizados hasta ahora, por tanto, es perfecto, dado que todos y cada uno presentan ciertas ventajas y desventajas, aunque no en idénticas proporciones. De ahí que si bien ninguno es totalmente objetivo o neutro, haya sin embargo considerables diferencias entre ellos, por lo que en función de los contextos y de los propósitos unos pueden ser más aconsejables que otros. Aquí me detendré en la exposición breve de algunos de ellos: los que gozan de una mayor aceptación entre los profesionales. Analizaré con más detenimiento el procedimiento de la evaluación del profesorado mediante la opinión de sus alumnos, al ser éste en la actualidad el más común dentro del mundo universitario y haber sido también utilizado desde finales del siglo XIX para evaluar la eficacia del profesorado de primaria y secundaria (Kratz, 1986). Además, este procedimiento es el que ha sido sometido a un mayor control científico, gracias al cual se ha podido ir respondiendo a las diversas críticas dirigidas tanto a la evaluación en general, como a este procedimiento en particular, a lo largo de la pasada centuria.

5.1. La observación

De entre todos los procedimientos, el primero, por derecho propio, es sin duda la *observación*. Es esta técnica la que ha servido históricamente de puente entre la evolución informal y formal a finales del siglo XIX y es ella la que, en la actualidad, sirve de denominador común para el resto de técnicas. Dentro de la observación, se pueden contar por cientos los distintos instrumentos ya elaborados -listas de comportamientos o escalas descriptivas, sistemas de categoría y de clasificación, informes, grabaciones, entre otros – (Good y Mulryan, 1990; Medley, 1972). Hasta ahora la mayoría de ellos han sido contruidos pensando en una enseñanza tipo como lo es la “lección magistral”, en la que el profesor es el protagonista activo y el alumno el receptor pasivo. Las *dimensiones* más frecuentemente evaluadas han sido:

- la dirección/control de la clase;
- su capacidad didáctica;
- sus características personales relacionadas con la enseñanza;
- su preparación académica y profesional;
- el interés por su trabajo;
- sus esfuerzos por mejorar;
- su capacidad para la cooperación.

Estos instrumentos han sido útiles para conocer el tipo de conductas desempeñado por el profesor en el aula, pero de poco o nada nos sirven si nuestro interés está centrado en los pensamientos o sentimientos del profesorado. Gracias a ellos, pues, podemos conocer el rendimiento del profesor pero no así el máximo de su competencia. En nuestros días, no obstante, la observación se ha extendido a todos los

miembros que componen el contexto específico de aprendizaje/enseñanza, a las interacciones entre ellos, y a los sistemas más amplios en los cuales se integran (familias, comunidades o sociedad). Por eso, las dimensiones a evaluar han ido aumentando, al igual que lo han hecho los aspectos a tener en consideración, como la participación del profesorado en las distintas actividades escolares, en la vida profesional y social o en el enriquecimiento de las actividades de la comunidad.

Un aspecto clave en los sistemas observacionales es el referido al *quién* realiza la observación. En la enseñanza no universitaria, esta labor ha recaído hasta muy recientemente en el personal directivo del centro, los inspectores o los administradores, que utilizaban sistemas de observación básicamente abiertos. Actualmente el predominio es de los sistemas de observación cerrados, aplicados por un personal mucho más variado (evaluadores, investigadores o formadores del profesorado, compañeros, directores de departamentos) y que, generalmente, ha recibido algún curso de formación para desempeñar esta labor específica.

La tendencia actual, que está ya marcando el futuro inmediato, es la de tratar de combinar múltiples fuentes de información, en este caso mediante las diversas técnicas de observación, sobre lo sucedido en el aula de un determinado profesor y ver su incidencia en el rendimiento y la calidad de ese producto de aprendizaje/enseñanza. Esta tendencia a la combinación se extiende más allá de las técnicas observacionales, con el fin de paliar las deficiencias consustanciales a las mismas. Ahora bien, es necesario tener en cuenta que, si como afirman Wise, Darling-Hammond, McLaughlin y Bernstein (1985), muy pocos de estos sistemas observacionales de evaluación del profesorado son realmente eficaces en Estados Unidos (ni siquiera para la resolución de problemas administrativos), entonces la combinación de resultados procedentes de diversas técnicas, por sí misma, no acabaría de resolver plenamente problemas fundamentales de la calidad docente del profesorado como el de la validez, el de la utilidad e incluso el del propio aprendizaje.

5.2. La autoevaluación

Lo que hemos aprendido a lo largo de más de un siglo de evaluación es que el profesorado, en general, acepta mejor la autoevaluación que la heteroevaluación y, dentro de ésta, la que está centrada en los aspectos formativos o de perfeccionamiento frente a la que tiene funciones administrativas. Aunque pueda parecer a primera vista que el concepto es diáfano y claro para todo el mundo, la realidad muestra que la autoevaluación se compone de muchos elementos (Bailey, 1981, nos habla de hasta 7 pasos en la realización de una autoevaluación), combinados de diferentes formas, aunque eso sí con un único propósito: la mejora de la propia docencia. La autoevaluación es un proceso continuo que implica que se han de identificar los comportamientos reales de uno mismo, se han de detectar los puntos fuertes y débiles y se han de intentar modificar esos puntos débiles en aras de la mejora docente (Airasian y Gullickson, 1997). Para ello se puede hacer uso de multitud de procedimientos que ofrecen una retroalimentación. Uno de los ejemplos más claros proviene de la utilización de la grabación de las propias clases. Otros, bastante utilizados, son las hojas de autoclasificación, el aprendizaje por observación de otros profesores considerados muy competentes, las opiniones de los alumnos, de los colegas o de los administrativos, la asesoría de expertos en asuntos de evaluación docente. Ahora bien, la autoevaluación en sí misma parece dejar considerables lagunas, debido básicamente a su falta de

objetividad, si no se complementa con otros procedimientos que faciliten la adopción de otras perspectivas capaces de equilibrar los sesgos derivados de la propia.

5.3. El portafolios/archivo o carpeta docente del profesor

Lo que uno, generalmente, puede encontrar en el portafolios/archivo o carpeta docente del profesor son planificaciones docentes, materiales para las clases, tipos de exámenes e informaciones provenientes de alumnos, colegas o autoridades académicas. Los evaluadores creen que estos documentos pueden resultar útiles para valorar el trabajo docente del profesorado, sobre todo cuando la finalidad es formativa, es decir, lo que se pretende es ayudar al profesor a mejorar su docencia y, por ende, su desarrollo profesional (Tigelaar, Dolmans, Wolfhagen, Van der Vleuten, 2004). La supervisión de estos materiales por el evaluador debe tener como finalidad el asesoramiento, evaluando positivamente todo lo que sea digno de reconocimiento (parte del portafolios donde se guarda el mejor trabajo del profesor) y realizando sugerencias sobre las posibles mejoras (parte del portafolios en donde se encuentra el peor trabajo del profesor), tal como señala Bird (1990). Este autor proporciona ejemplos ilustrativos sobre las distintas formas (habla de 9) de organizar estos archivos con el fin de que los materiales no constituyan una documentación caótica, sino una información muy diversa pero organizada según unas categorías significativas.

Cano (2005) considera 17 categorías:

- *apartado inicial (afiliación, currículo);*
- *índice (sumario, mapas conceptuales);*
- *justificación e interrelación de los organizadores (hilo conductor, enlaces);*
- *referentes teóricos (coherencia entre teoría y práctica);*
- *filosofía docente y objetivos (la propia concepción docente y su materialización en fines concretos);*
- *diarios, textos paralelos, registros anecdóticos (diarios abiertos, semiestructurados, estructurados; textos que describen y muestran la reflexión o valoración; observaciones ilustrativas);*
- *aspectos éticos relativos a la enseñanza (compromiso, solidaridad);*
- *materiales curriculares (planificación e intervención);*
- *borradores (que muestren la evolución del ciclo docente);*
- *notas de los mentores;*
- *observaciones de la propia clase (para ver el grado de materialización de lo propuesto);*

- *observaciones de las clases de los demás (modelos);*
- *la satisfacción de los alumnos;*
- *los aprendizajes de los estudiantes (teniendo en cuenta varios aspectos de su rendimiento, además de las notas);*
- *comisiones o comités (labores de gestión);*
- *autoevaluación (profesional responsable);*
- *aprendizajes propios (desde las más diversas perspectivas: compañeros, materiales o diversos contextos).*

El archivo, organizado según estas categorías u otras parecidas, nos ofrecería una muestra ilustrativa de la postura de ese profesor ante las principales cuestiones a las que ha de hacer frente en un contexto de aprendizaje/enseñanza: sus relaciones con los alumnos, con los compañeros, con las autoridades o con los padres de sus alumnos. Si la finalidad cambiase (simple evaluación cosmética o ritual), el archivo se constituiría en un mero ejercicio de acumulación de papel o incluso en arma contraproducente para el propio profesor, pues todo su trabajo sería una pérdida de tiempo (Brown y Wolfe-Quintero, 1997). Ahora bien, además de la finalidad formativa o del desarrollo profesional –la principal y la más aceptada en la comunidad científica–, las carpetas docentes también pueden ser utilizadas para su función sumativa, aunque aquí surgen muchas más dudas en torno a su validez. Las comparaciones entre carpetas docentes son difíciles, debido a la libertad de los profesores para elaborar su portafolios, a la par que resulta complicado evitar los posibles sesgos derivados de la preponderancia dada a la subjetividad.

Este procedimiento, que como todos los demás cuenta con sus ventajas y desventajas –aunque en modo alguno en igual grado–, está siendo utilizado en la actualidad en los distintos niveles educativos: primaria, secundaria y universidad (Black, Daiker, Sommers y Stygall, 1994; Davis y Swift, 1995; Fernández, 2004; Moss, Schutz y Collins, 1998; Shores y Grace, 2004) y empleado en muy diversos países (Crispín y Caudillo, 1998; Ministerio de Educación de Chile, 2003). Es muy probable que su base constructivista –enfoque predominante en la psicología actual– y su encuadre dentro de las técnicas cualitativas –en alza en estos momentos en diversas disciplinas– estén sirviendo de lanzaderas para su actual extensión. A ello ha de añadirse la incorporación de las nuevas tecnologías en la elaboración de este procedimiento: la carpeta docente electrónica (Barrett, 2000; Hauge, 2006; Heath, 2002).

5.4. Las opiniones de los alumnos

Con alta probabilidad, uno de los trabajos más completos y rigurosos sobre la evaluación del profesorado a través de las opiniones de sus alumnos, en el ámbito universitario, es el de Marsh (1987). De ahí que lo tome como fuente de documentación básica para este apartado. Se puede afirmar que este procedimiento de evaluación se inicia en Estados Unidos en torno a 1920 en algunas de sus más famosas universidades y que desde entonces no ha dejado de extenderse hasta alcanzar, en nuestros días, la práctica totalidad de sus universidades. Además, no es ya sólo Estados Unidos sino la

mayoría de los así llamados países desarrollados los que lo han comenzado a introducir en sus universidades. Incluso países en vías de desarrollo también se están haciendo eco de este sistema de evaluación para sus profesores universitarios. Como cabe imaginar, este movimiento en expansión ha producido tal cantidad de instrumentos de evaluación que ocuparían varios volúmenes si tratásemos de poner unos al lado de otros. Gran parte de los mismos goza de buenas propiedades psicométricas, lo que facilita las posibles elecciones para aquellos que no se quieran molestar en elaborar uno propio (Centra ya nos ofrece, a finales de la década de los 70, un Apéndice dedicado a algunos de los más conocidos: Centra, 1979).

En general, se puede decir que hasta ahora los resultados obtenidos por estos medios han servido a algunos de los siguientes propósitos:

- *mejorar la calidad de enseñanza del profesor*, gracias a la devolución de la información proporcionada por sus alumnos;
- conocer la calidad docente de un profesor a fin de poder *tomar decisiones de tipo administrativo* (promoción, contrato fijo, rescisión del contrato);
- conocer la calidad del profesor y del curso por parte de los alumnos interesados, y así poder decidirse fundamentadamente a la hora de sus *elecciones sobre profesores/materias*;
- conocer la *calidad del curso a fin de poder mejorarlo*;
- contar con *radiografías que permitan la investigación* sobre la calidad o excelencia de la enseñanza.

Una de las cuestiones lógicas que más corpus bibliográfico ha producido es la referida a la *dimensionalidad* de la calidad docente, materializada a través de los distintos factores de los diversos instrumentos de evaluación (Abrami, 1985). El abanico va desde *1 ó 2* dimensiones (Frey, 1978) hasta las *19* recogidas por *Feldman* (1976): estimulación del interés por la materia, entusiasmo, conocimiento, amplitud de miras, preparación de las clases, claridad en la exposición, disponibilidad/accesibilidad, respeto por los alumnos, gestión de la clase, imparcialidad evaluadora, etcétera. Marsh (1984, 1987), va a defender *9 dimensiones* que son las que precisamente valora su famoso *SEEQ* (Students' Evaluations of Educational Quality). Creo que se ha sobrevalorado este asunto, pues el número de dimensiones debería estar enmarcado dentro de algún tipo de teoría y no ser meramente el resultado de ir sumando elementos, sometiéndolos después a los pertinentes análisis factoriales, bien de tipo exploratorio o incluso confirmatorio. Esta es la crítica que en su momento sacamos a la luz cuando publicamos nuestros propios instrumentos de evaluación desarrollados dentro de la Universidad Complutense (Angulo, Fernández y Martínez Arias, 1987; Fernández y Mateo, 1992; Mateo y Fernández, 1993).

En la actualidad, tras este cúmulo de investigaciones utilizando como fuente de información las opiniones de los alumnos, sobre todo en el ámbito universitario, cabría asumir los siguientes *asertos*, sabiendo que están considerablemente bien fundamentados:

- Parece que se impone la *multidimensionalidad* a la hora de entender operativamente la calidad docente.
- Buena parte de los instrumentos de valoración publicados muestran *índices de fiabilidad satisfactorios*.
- Estos instrumentos resultan relativamente *válidos*, en tanto los resultados obtenidos con ellos no parecen estar influidos significativamente por otras variables distintas de las que se pretende valorar.
- Los resultados recogidos mediante estos instrumentos suelen resultar *útiles para los alumnos, los profesores y las autoridades administrativas*, sobre todo por lo que respecta a la mejora de la calidad docente (Feldman, 1976; Marsh, 1983).

Esto en modo alguno implica que se hayan de considerar los únicos medios válidos para evaluar al profesorado o que no muestren ciertas *deficiencias* (Wilson, 1988). Entre éstas cabría destacar:

- El hecho de que este tipo de evaluaciones pueden ser susceptibles, en algún caso, del *efecto halo*, o del *efecto del Dr. Fox*: los *profesores seductores*, gracias a su entusiasmo o expresividad, influyen en la valoración –positiva– de sus alumnos, con cierta independencia del contenido impartido (Meier y Feldhusen, 1979; Perry, Abrami y Leventhal, 1979).
- La constatación de que la *correlación con otros criterios de calidad docente* no es tan alta como cabría esperar, a raíz de los análisis de las propiedades formales de los distintos instrumentos de evaluación.
- La comprobación de que estas evaluaciones pueden verse afectadas, en un cierto grado, por *variables extrañas* (benevolencia en las calificaciones de los alumnos, etcétera).
- La verificación de un cierto *rechazo del profesorado* cuando se han querido utilizar los resultados de las evaluaciones para ejecutar decisiones administrativas.

Teniendo en cuenta *los puntos positivos y negativos*, resulta difícil dudar de su utilidad al menos con respecto a estas facetas:

- Como *retroalimentación* para el profesorado, gracias a la cual éste puede ir mejorando su docencia.
- Como base de *elección de las clases* por parte del alumnado. Estos datos gozan de una mayor fiabilidad y validez que la que pueden proporcionarse los alumnos por sí mismos a través de amigos o compañeros.

- Como *información necesaria* a la hora de investigar la calidad docente.

Tal vez esto explique el porqué este procedimiento evaluador sea uno de los más utilizados y de mayor calidad, al haber superado un gran número de pruebas empíricas tanto en torno a su fiabilidad como a su validez. Ahora bien, dado que también se han puesto de manifiesto ciertas deficiencias, parece aconsejable su utilización junto con otros procedimientos como pueden ser: la evaluación de los pares, la autoevaluación o la evaluación de las propias autoridades académicas.

5.5. Otros procedimientos

Lo expuesto hasta aquí sólo ha de considerarse una muestra ilustrativa de las distintas formas de acercarse a la evaluación del profesorado. Sin duda alguna, hoy en día contamos con bastantes más procedimientos, como pueden ser las entrevistas (estructuradas, semiestructuradas, poco estructuradas o informales), las valoraciones de los inspectores, el director del centro u otras autoridades académicas, las evaluaciones realizadas por los propios compañeros (del centro o de fuera del centro), la comprobación de los logros académicos de sus alumnos, la satisfacción de los padres o diversas medidas indirectas de la competencia del profesor (actividades académicas fuera del aula, características psicológicas del profesor, intereses y actividades no profesionales del profesor).

Dada la abundancia de métodos o formas existentes a la hora de evaluar la calidad docente del profesorado (Millman, 1981; Millman y Darling-Hammond, 1990), tal vez la cuestión fundamental no sea la elección de un procedimiento frente a otro. Aquí surge lo que considero un falso problema pero que está produciendo ríos de tinta: la excesiva confianza en la bondad de un procedimiento frente a los demás (cuantitativos frente a cualitativos; cuestionarios frente a entrevistas; opiniones de los alumnos frente a las de los pares, etcétera). Me parece que la verdadera cuestión reside en definir con claridad qué es lo que se quiere evaluar (definirlo lo más operativamente posible) para después encontrar al evaluador debidamente preparado, capaz de seleccionar o diseñar un procedimiento adecuado a tal fin. Los análisis estadísticos son sin duda de gran ayuda; la recogida de la información de una u otra forma es absolutamente imprescindible, pero todo ello ha de ser correctamente interpretado (sólo lo suelen hacer bien las mentes especialmente preparadas) para que, si procede, se puedan tomar las correspondientes decisiones, bien de tipo formativo o sumativo. En mi experiencia como evaluador, la variable que ha explicado mayor proporción de varianza sobre la bondad de un determinado proceso valorativo ha sido la referida a la formación específica del evaluador: si no selecciona bien los medios en función de los fines o si no es capaz de elaborar el medio más adecuado para un contexto determinado, el resto de variables no suelen paliar esta deficiencia esencial. Si, en cambio, la mente del evaluador funciona adecuadamente, la diversidad de procedimientos desarrollados para un mismo fin le proporcionarán prácticamente la misma información relevante, aunque con costes diferentes: lo que le digan los alumnos y lo que se le ofrezca mediante la observación será en esencia lo mismo. Lo dicho en modo alguno ha de interpretarse como si todos los procedimientos fuesen iguales: me estoy refiriendo, como resulta lógico, a los buenos medios, a aquéllos que nos aseguran fiable y válidamente la información recogida. Sólo descartados los malos medios (que en ocasiones son los más abundantes), de uno u otro signo (opiniones de los alumnos, de los compañeros, de los

inspectores, etcétera), tendría pleno sentido lo dicho anteriormente. A su vez, esta mente cultivada (adecuadamente preparada) del evaluador sabrá discernir el medio más apropiado cuando se trate de valorar aspectos ya muy específicos. Así, por ejemplo, las estrategias cognoscitivas que utiliza un profesor a la hora de preparar una buena clase sólo es posible conocerlas directamente a través de procedimientos centrados en lo que nos cuente el propio profesor, no en lo que nos puedan decir los pares, los alumnos, los padres, los inspectores o las diversas autoridades.

6. Comentarios

Difícilmente puede llegar a buen puerto la evaluación del profesorado si ésta no se encuadra dentro de una cultura de la evaluación que tenga como horizonte la profesionalización del profesorado. Dentro de este marco, la evaluación aparece como un mecanismo imprescindible para la mejora de la calidad docente del profesor y del propio sistema educativo.

En nuestros días, esta evaluación ha de ser necesariamente muy abarcadora. Gracias a ella se han de valorar, por supuesto, los conocimientos del profesorado, pero no en menor grado sus capacidades, sus competencias, sus sistemas de valores o sus actitudes. Además, la evaluación ha de concebirse como un proceso que se prolongará a lo largo de toda la carrera profesional, ofreciéndonos información relevante sobre todos y cada uno de los momentos clave de la misma: titulación, programas de formación, selección, primeros años de ejercicio profesional, etcétera.

Esta evaluación se puede concretar a través de distintos procedimientos (observación, autoevaluación, portafolios, opiniones de los alumnos, entre otros), cada uno de ellos con sus puntos fuertes y débiles, aunque el juicio global no sea en modo alguno igual para todos ellos. Unos son más idóneos que otros, aunque ninguno de ellos sea del todo perfecto.

Dado que en el futuro tendremos, cada vez más, a nuestra disposición un mayor número de técnicas evaluadoras, parece adecuada la exigencia tanto de un evaluador debidamente preparado (no cabe sustituirlo por un profesional que muestra cierta afición o interés en el campo de la evaluación) como de una tecnología (muy frecuentemente asentada en algún conocimiento de tipo estadístico) capaz de facilitarle una interpretación verdaderamente significativa de todo el cúmulo de información recogida.

CAPÍTULO CUARTO

El modelo de evaluación circular

1. Introducción

Dentro de los múltiples procedimientos con los que hoy contamos para llevar a cabo la evaluación del profesorado, voy a exponer a continuación el que se ha desarrollado en la Universidad Complutense de Madrid a lo largo de las dos últimas décadas del siglo XX (Angulo, Fernández y Martínez, 1987; Fernández, 1988). Una síntesis del mismo puede verse en la revista de la Organización para la Cooperación y el Desarrollo Económico (OCDE), en francés y en inglés, o en uno de los libros publicados por esta misma institución (Fernández, 1992, 1997; Fernández y Mateo, 1994 -se cita aquí la versión inglesa-). Uno de los instrumentos que he desarrollado con tal fin aparece recogido en el Handbook de Lester y Bishop (2000), libro en el que se pretenden recoger más de un centenar de instrumentos que supuestamente, según los autores, han aportado alguna novedad en el ámbito de las ciencias sociales y muy concretamente en el terreno de la educación.

Si se me pidiera que resumiese lo más sintéticamente posible en qué consiste el *modelo de evaluación circular*, diría que consiste en que *todos y cada uno de los agentes o miembros que participan y constituyen un determinado sistema –en este caso el sistema educativo– pueden y deben ser evaluados por el resto de los agentes, según sus capacidades y competencias específicas, a fin de poder cruzar la información recogida desde estas diferentes perspectivas y así poder tomar decisiones bien fundamentadas científicamente, bien sean éstas de tipo formativo o sumativo.*

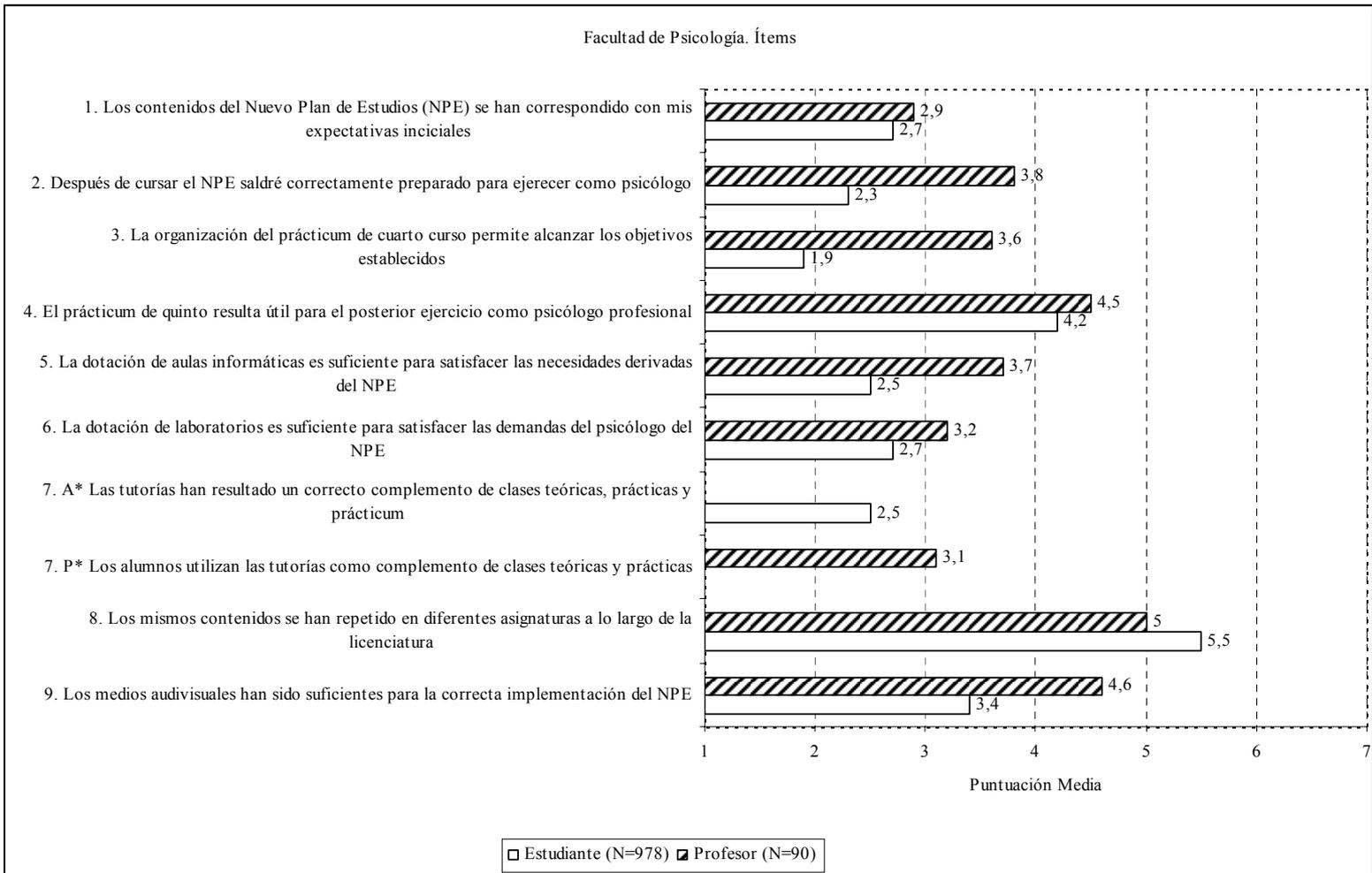
El *alumno* podrá y deberá evaluar aquellos comportamientos del profesor (se incluyen aquí sus capacidades, competencias o sistemas de valores) que incidan directa o indirectamente en la calidad docente que recibe, aunque no, por ejemplo, su competencia investigadora, así como el *profesor* podrá y deberá evaluar el comportamiento y rendimiento de todos y cada uno de sus alumnos lo más rigurosa y objetivamente posible (también aquí se incluyen las capacidades, competencias y sistemas de valores). *El personal de administración y servicios (PAS)* evalúa, al menos informalmente (se le debiera dar la oportunidad de que lo hiciera formalmente), los comportamientos de alumnos y profesores (dedicación a los estudios, trato humano, etcétera), así como estos dos colectivos valoran la actuación del PAS (grado y tipo de ayuda que proporciona o problemas que puede causar). A su vez, las *autoridades académicas y políticas* evalúan formalmente las actividades de alumnos y profesores (rendimiento de ambos colectivos), así como estos dos grupos evalúan la actuación de dichas autoridades (las normativas legales que establecen, el distanciamiento o cercanía, etcétera).

Una de las características más destacadas de este modelo de evaluación circular es la que se refiere al *cruce de los datos* obtenidos de los diferentes agentes. Gracias a este cruce es posible conseguir una radiografía mucho más rica en matices que la que se lograría mediante evaluaciones –las más frecuentes– en las que siempre aparecen unos como evaluados y otros como evaluadores (éstos suelen ser los profesores y aquéllos los alumnos), o aquellas otras –poco frecuentes– en las que a lo más que se llega es a yuxtaponer las distintas perspectivas (por un lado se analizan los resultados de un tipo de miembros –por ejemplo, los profesores–, para después pasar al análisis del otro u otros tipos –por ejemplo, los alumnos, el PAS o las autoridades académicas–).

Tratemos de verificar lo dicho con un ejemplo real. Me refiero ahora a la *evaluación de un plan de estudios de la Titulación de Psicología*. Para ello elaboramos dos instrumentos

de valoración *ad hoc*: uno para que lo respondieran los *alumnos* y otro los *profesores*. Una mínima parte de los resultados obtenidos se recogen en la Tabla 2.

Tabla 2
Comparación entre las evaluaciones realizadas por estudiantes y profesores del plan de estudios de Psicología (Fernández, 2001)



A*: Ítem de alumnos; P*: Ítem de profesores

Desde la perspectiva más clásica, muy probablemente sólo se hubiese recogido la opinión del profesorado, al pensar que sólo este colectivo estaría capacitado para emitir juicios certeros en torno a la realidad a la que se refieren los ítems. Desde una perspectiva hoy por hoy minoritaria, al menos dentro del ámbito educativo, se hubiese preguntado a profesores y alumnos por separado, sin cruzar los datos y analizar el significado nuevo que ofrece el cruce. Desde el modelo de evaluación circular se tienen en cuenta, primeramente, las opiniones de cada colectivo por separado, para después compararlas de forma cruzada. Gracias a este enfoque constatamos, cuando ya se realiza la comparación cruzada (los análisis de cada grupo por separado debieran arrojar las mismas valoraciones desde los distintos enfoques) lo siguiente:

- que hay mucha *coincidencia en ciertos ítems negativamente valorados* (sirva el ítem 1 como ejemplo), o *relativamente bien valorados* (ítem 4), reforzando así considerablemente el valor y el significado de los mismos;
- que hay *discrepancia en otros* (el caso más patente es el ítem 3), lo que debiera ser un estímulo muy poderoso para una reflexión ulterior de cada grupo y de las autoridades académicas a fin de que todos ellos “aprendieran” que una “realidad” manifiesta facetas muy distintas en función de la perspectiva que se adopta de partida;
- que hay asuntos tremendamente importantes que han de ser modificados, sobre los cuales se da un *consenso pleno* entre los dos protagonistas del sistema educativo: profesores y alumnos (ítem 8), por lo que se debieran tomar decisiones de inmediato, pese a los esfuerzos que tal tarea pueda suponer para el profesorado y autoridades académicas;
- que hay total *acuerdo entre estos dos colectivos a la hora de vislumbrar la solución* (prácticum de quinto –ítem 3-) para aquello que no funciona (prácticum de cuarto –ítem 4-), aun constatando diferencias claras entre alumnos y profesores con respecto a lo que funciona inadecuadamente;
- que se requiere una *reflexión especial* en torno al ítem 7 (referido a la complementariedad de las tutorías), pues si bien hay coincidencia en la valoración negativa por parte de alumnos y profesores, sin embargo, las causas de esta valoración difieren en ambos colectivos: los alumnos las atribuyen a los profesores y éstos a aquéllos.

Un segundo ejemplo, tomado también de esta misma investigación (véase Tabla 3), nos puede ser útil para dar respuesta clara a una cuestión muy frecuente dentro del mundo académico: ¿los alumnos universitarios, *aprendices*, realmente están capacitados para saber qué es lo que se les debe enseñar dentro de una Titulación específica?

Tabla 3
Selección de las asignaturas de primer curso en función de su relevancia (las más relevantes)

ALUMNOS	%	PROFESORES	%
		Métodos y diseños de investigación en psicología I	76

Psicología del aprendizaje	76	Psicología del aprendizaje	74
Fundamentos de neurociencia	68	Estadística aplicada a la psicología I	73
Estadística aplicada a la psicología I	59	Introducción a la psicología	69
Introducción a la psicología	55	Fundamentos de neurociencia	58
Estadística aplicada a la psicología II	53	Estadística aplicada a la psicología II	52

De las *10 asignaturas* que debían evaluar como *más relevantes* en el primer curso, los alumnos señalan 5 y los profesores 6. Hay coincidencia en considerar a 5 como muy relevantes y, además, hay igualmente coincidencia con respecto a las 4 que no aparecen en la tabla, puesto que son consideradas como *menos o poco relevantes* tanto por los profesores como por los alumnos. ¿Cómo se puede explicar esta considerable coincidencia si los alumnos estuvieran incapacitados para evaluar? ¿Esta concomitancia no nos está indicando el camino a seguir a la hora de plantearse la mejora de dicho plan de estudios? Con respecto a la única asignatura en la que aparecen discrepancias, la información proporcionada por otros ítems nos descubre que en la valoración negativa de los alumnos había influido considerablemente su profesorado. Es muy probable, pues, que si se subsanase este problema, la coincidencia fuese todavía mayor (total), aunque siempre hay que tener en cuenta que el ordenamiento de las asignaturas en función del porcentaje de alumnos y profesores nos obliga a realizar las lógicas y coherentes matizaciones que cabría esperar de dos colectivos tan bien diferenciados.

Tras este primer acercamiento a lo que entiendo por modelo de evaluación circular, ejemplificado mediante datos reales, seguiré ahora con una exposición pormenorizada que se centra básicamente en los dos componentes clave de cualquier proceso de aprendizaje/enseñanza: los alumnos y los profesores, tal cual se especificó en el capítulo primero (véase Figura 1).

2. Las opiniones de los alumnos

Como en su momento se indicó (capítulo tercero), las opiniones de los alumnos han sido utilizadas en la enseñanza no universitaria para evaluar a sus profesores. Mucho más extendida se halla la práctica de solicitar la opinión del alumnado para la evaluación del profesorado en el ámbito universitario. De hecho, en los Estados Unidos, este tipo de evaluación se remonta a los inicios del siglo XX. Hoy en día es práctica generalizada en las universidades estadounidenses, en buena parte de los países europeos y en bastantes universidades iberoamericanas. En general, pues, cabe hablar de un tipo de evaluación rutinaria en los países desarrollados y que incluso ya se comienza a introducir en los países en vías de desarrollo (Dawoud, 1983; Miron, 1988; Murray, 1984; Rushton & Murray, 1985; Watkins, Marsh, & Young, 1987).

2.1. Las variables relevantes frente a las irrelevantes

Parece lógico que si se quieren recoger las opiniones de los alumnos, como procedimiento de evaluación de la calidad de sus profesores, se piense ante todo en hacerlo de forma válida y fiable. Esto requiere que se tengan en cuenta las aportaciones realizadas por los distintos investigadores a lo largo de todo el siglo XX. En nuestro caso, esta revisión bibliográfica supuso descubrir una serie de variables que sí parecían incidir en la calidad docente, junto con otras que habían mostrado su escasa o nula relevancia a la hora de comprender los determinantes de referida calidad o excelencia. Entre las primeras, nos encontramos con: *los contenidos del curso, la*

coherencia/claridad en las explicaciones, el nivel de conocimientos del profesor, la eficacia en la presentación/organización de los contenidos, la utilidad del trabajo adicional a la clase, el interés despertado en el alumno por la asignatura, la tolerancia del profesor hacia los distintos enfoques, el interés del profesor por la asignatura, la accesibilidad del profesor, la interacción profesor-grupo o la estimulación de la participación, por citar algunas de las más relevantes. Entre las escasamente relevantes aparecieron: *el sexo del profesor, su rango académico, el número de años docentes o la dificultad de la asignatura*, por señalar algunas de las más estudiadas.

Esto nos permitió elaborar un instrumento de valoración, denominado *Dimensiones del Profesorado Universitario –DPU-* (ver el apartado de Anexos: Anexo 1), en el que se mezclan a propósito las variables relevantes e irrelevantes. Su posible utilidad radica en que constituye una síntesis de multitud de estudios empíricos internacionales en los cuales se han ido explorando todas y cada una de las variables que el instrumento recoge. Nuestro primer objetivo fue comprobar hasta qué punto los resultados obtenidos en otros contextos –estadounidenses principalmente- se podían replicar en España, más concretamente en la Universidad Complutense. Aquí empezamos a constatar algo que resulta esencial en el ámbito de la evaluación: siempre hay que poner a prueba tanto los instrumentos como los resultados que provienen de otros contextos. Esta postura se contrapone a lo que, por desgracia, suele ser bastante frecuente (“ley del mínimo esfuerzo”): traducir o extrapolar sin más lo que ya han realizado los demás.

El trabajo empírico (aplicación del DPU), realizado en la Complutense con un grupo de profesores y otro de alumnos (modelo de evaluación circular), nos permitió llegar a las siguientes conclusiones:

- *Coincidencia de los resultados recogidos en las bases de datos internacionales y los obtenidos en España.* Ratificación, pues, mediante *réplica*, de los conocimientos sobre las variables a tener en cuenta en los análisis sobre la calidad docente.
- *Coincidencia de profesores y alumnos* tanto en las principales variables consideradas relevantes como en aquellas otras valoradas como irrelevantes.
- *Discrepancias entre profesores y alumnos* en ciertas variables (más bien pocas) que, según se había puesto de manifiesto en la bibliografía internacional, afectaban más a unos que a otros, bien perteneciesen aquéllas al grupo de las relevantes o al de las menos relevantes: los alumnos, más que los profesores, consideran relevantes (medias más altas), por ejemplo, la ecuanimidad/justicia en las calificaciones, la utilidad/aplicabilidad de los contenidos, la bibliografía recomendada o el trabajo adicional a la clase. Por el contrario, los profesores consideran más relevante que los alumnos el currículum investigador a la hora de hablar de la excelencia docente.

Con esto en mente, ya se podía dar el paso siguiente en el proceso evaluador: la elaboración de un nuevo instrumento de evaluación que incluyera todas las variables relevantes (puestas así de manifiesto tanto por la bibliografía internacional como por los profesores y alumnos de nuestra universidad), sin incluir ninguna de las irrelevantes. El

producto fue el *Cuestionario de Evaluación de los Profesores de la Universidad Complutense (CEPUC, en castellano y CUTEQ, en la versión inglesa)*.

2.2. Elección o elaboración de instrumentos

El siempre obligado repaso al corpus bibliográfico internacional antes de comenzar cualquier proceso evaluador nos posibilita conocer lo ya elaborado (a fin de no estar “descubriendo continuamente el mediterráneo”), además de los posibles resultados de investigaciones realizadas para comprobar hasta qué punto es posible su utilización en un determinado contexto, como puede ser el nuestro. Esto fue lo que hicimos teniendo en cuenta un objetivo concreto: encontrar posibles instrumentos de evaluación que cumplieran el requisito de incorporar las variables relevantes, dejando fuera las irrelevantes. El más conocido internacionalmente, el que había recibido hasta la fecha más apoyos empíricos, el que gozaba de una mayor y mejor fundamentación era el de Marsh (Marsh, 1984, 1987). Ésta parecía a todas luces la opción ideal. Sin embargo, no fue la elegida por nosotros, pues ya contábamos con el estudio antes comentado (el del DPU) y pensamos que merecía la pena elaborar un nuevo instrumento que, inspirado por supuesto en el de Marsh, recogiese algunos de los matices específicos de nuestra situación universitaria, dejando de lado aquellos aspectos que, siendo esenciales para las universidades australianas o estadounidenses, lo eran menos para las nuestras.

El instrumento original, el CEPUC (Cuestionario de Evaluación del Profesorado Universitario de la Complutense), consta de 39 ítems que tratan de desglosar y materializar las variables relevantes encontradas (véase Angulo et al., 1987 para un análisis más detallado del ofrecido a continuación). En cuanto a sus propiedades formales, es pertinente indicar que goza de una alta fiabilidad, en tanto consistencia interna ($\alpha = .97$), pero deja margen para la mejora en cuanto a la validez de constructo, pues los factores no nos dan cuenta de toda la varianza que sería de desear, aunque en modo alguno puede considerarse insatisfactoria (en torno al 68%) y, además, de los 5 factores solamente el primero es auténticamente robusto (sólo él da cuenta de un 53% de la varianza). Este primer factor consta de 25 ítems y resume en él dos amplios conglomerados (el de la competencia docente del profesor y el de sus habilidades para motivar al alumnado), así detectados tras los correspondientes análisis de conglomerados (*cluster analysis*), por lo que recibió la denominación de Competencia docente y motivacional.

Además de por *razones técnicas* (mejora de la validez de constructo en lo posible), otras *razones de tipo práctico* nos indujeron a la necesidad de una reducción de los ítems, pues resultaba enormemente cansado y aburrido para los alumnos tener que repetir para cada uno de sus profesores (un promedio de 5) la valoración, según una escala tipo Likert (de 1 a 7), de todos y cada uno de los 39 ítems. En principio, seleccionamos de entre los 39 los 25 que, tanto desde un punto de vista teórico (mejor materialización de las variables relevantes), como empírico (mostraban una alta saturación en el primer o segundo factor de los 14 análisis factoriales realizados y una correlación superior a .40 en todos o la mayor parte de los 14 análisis de conglomerados), podían materializar mejor, bien la competencia docente del profesor, o bien sus habilidades motivacionales y de interacción. Tras un estudio piloto con esta versión reducida del CEPUC (*CEPUC-R, en castellano, y CUTEQ-R, en la versión inglesa*) en que aparecieron los dos factores previstos: el de la competencia docente y el

de las habilidades motivacionales y de interacción, se pasó a ponerlo a prueba en todos los Centros de la Complutense (un total de 36.589 estudiantes), suprimiendo 2 ítems cuyos contenidos podían trascender las capacidades evaluadoras de los estudiantes, según manifestaron las autoridades de esta universidad. Esta versión abreviada de 22 ítems, más uno de carácter general que ejercía de criterio, volvió a mostrar una alta consistencia interna ($\alpha=.98$) y los dos factores previstos que explicaban en torno al 71% de varianza. Los detalles más técnicos (análisis estadísticos) y una visión más pormenorizada pueden consultarse en Fernández y Mateo (1992) y Mateo y Fernández (1992).

La versión que más se ha utilizado es una intermedia que consta de 29 ítems (ver apartado de Anexos: Anexo 2), pues parecía que, además de las dos dimensiones apoyadas empíricamente en los análisis del CEPUC-R, mediante análisis factoriales exploratorios y confirmatorios, los análisis de conglomerados realizados con la versión inicial nos habían mostrado un par de conglomerados omnipresentes que parecía conveniente tener en cuenta a la hora de comprender la calidad docente: la ecuanimidad evaluadora y los requisitos docentes. Además, para facilitar al máximo la comprensión por parte del profesorado de los resultados obtenidos tras la evaluación del alumnado, los 29 ítems se agrupaban en una serie de *núcleos teóricos* (ver cuadro adjunto) que cuentan con el correspondiente respaldo empírico (los resultados de los análisis de correlación canónica y de regresión simple y múltiple aparecen recogidos en Mateo y Fernández, 1993).

Cuadro 5

Núcleos teóricos en que se agrupan los 29 ítems del CEPUC-intermedio

Núcleos teóricos	Ítems (redacción abreviada)
Estilo Docente	1. Su forma de enseñar facilita la comprensión de la materia
	4. Despierta interés en el alumno por la asignatura
	10. Manifiesta entusiasmo por los contenidos explicados
	13. Su explicación mantuvo mi atención en las clases
	15. Atrae la atención del alumno hacia la materia explicada
Estructuración de contenidos	7. Alude a las implicaciones de los distintos enfoques
	8. Establece una adecuada interrelación entre los temas
	9. Refleja la utilidad de los contenidos expuestos
Claridad expositiva	12. Responde a las preguntas con precisión
	14. Explica de forma coherente los contenidos de cada tema
	16. Explica con claridad
Competencia académica	18. Se muestra competente en los contenidos que imparte
	19. Manifiesta una adecuada preparación previa a la clase
Habilidades motivacionales y de interacción	2. Muestra interés por las inquietudes de cada alumno
	3. Motiva a preguntar sobre cualquier duda
	5. Se muestra accesible en su relación con los alumnos
	6. Facilita la expresión de las ideas propias de los alumnos
	17. Estimula la participación del alumno en la clase
Requisitos docentes	20. Es constante en la asistencia a clase
	21. Lo explicado se ajusta a los objetivos del programa
	25. La bibliografía recomendada sirve para preparar la materia
	22. El procedimiento de examen permite al alumno reflejar sus conocimientos

Ecuanimidad evaluadora	23. Examina sobre contenidos considerados como fundamentales
	24. Las notas reflejan el nivel de conocimientos mostrados por los alumnos
Criterios generales	11. Gracias a lo aprendido siento interés por esta materia
	26. Me gustaría recibir de nuevo otra asignatura con este profesor
	27. La eficacia docente de este profesor
	28. Calificación comparativa del profesor con los restantes del curso
	29. Calificación comparativa de esta asignatura con las restantes del curso

Aquí parece obligado que tratemos de dar algún tipo de solución a una de las cuestiones que más parece haber preocupado a la comunidad de los expertos en evaluación durante la segunda mitad de la pasada centuria, al menos a aquéllos que tratan de recoger las opiniones de los alumnos mediante cuestionarios. Cuando se habla de calidad docente, ¿cuántas son las dimensiones que mejor la materializan? La opinión de los distintos expertos varía considerablemente. Unos consideran que con **2** sería suficiente. De una u otra forma se están refiriendo a los dos factores que hemos visto que aparecían tras la aplicación de nuestro instrumento de evaluación. Otros, como es el caso más conocido de Marsh (1987), nos hablan de **9**, que son las que mide precisamente su cuestionario. Un tercer grupo nos habla de dimensiones que o bien están entre los dos números señalados o incluso sobrepasan al mayor.

Comenzado ya el siglo XXI, creo que esta no es la cuestión más relevante, sobre todo cuando se está pensando en la técnica del análisis factorial como el medio más adecuado para su apoyo empírico. Entiendo que debería ser la teoría subyacente sobre la calidad docente el asunto más importante a tratar. Si contásemos con esa teoría, el acuerdo en torno a las dimensiones quedaría reducido a una cuestión más bien de tipo técnico que, si bien en modo alguno carecería de importancia, quedaría sin embargo relegada a un segundo plano. Muy probablemente la teoría se constituiría en un marco lo suficientemente potente como para que la variación en torno al número de dimensiones se redujese considerablemente.

2.3. La devolución de la información

En vez de elegir un instrumento que ya gozaba de buenas propiedades formales, como era el de Marsh, optamos, como acabamos de ver, por la elaboración de uno propio. Esto, como resulta lógico, tiene sus ventajas y desventajas. Entre las primeras cabría hablar de la mejor adaptación a las especificidades contextuales, pero entre las desventajas, nos encontraríamos con la dificultad de llevar a cabo comparaciones de los resultados obtenidos en muy diversos ambientes. En ciencia, a nadie le puede caber el honor y la satisfacción de ser totalmente perfecto, aunque unos autores se acercan más claramente que otros al ideal de la perfección, cuando logran minimizar las desventajas y consiguen maximizar las ventajas. El hecho de que nadie sea perfecto, en modo alguno quiere decir que todos los investigadores o todos los métodos o procedimientos, dentro del ámbito de la evaluación, sean equivalentes. Para que el lector pueda juzgar por sí mismo, y así tomar la decisión que le parezca más oportuna sobre la conveniencia de elegir o elaborar instrumentos, seguimos con la exposición de otra faceta importante del modelo de evaluación circular: su peculiar formato a la hora de presentar al profesorado los resultados de sus evaluaciones. Un vez más, hemos preferido elaborar algo propio a seleccionar lo realizado por los demás, siendo muy conscientes, de nuevo, de los pros y los contras de esta decisión (Fernández, 1992).

Como se puede comprobar, analizando el apartado de Anexos (Anexo 4.1), nos pareció oportuno que el formato de devolución de la información para cada profesor (más de 4.000 en el caso de la Universidad Complutense) comenzase con tres hojas cuyo objetivo es *facilitar al máximo la correcta interpretación de los datos referidos a su docencia*. En ellas se les indica que se va a encontrar con:

- Una *ficha técnica*, donde puede ver descritos todos los aspectos relevantes de la evaluación (centros, profesores, asignaturas evaluadas, alumnos evaluadores, evaluaciones realizadas); los datos más significativos del instrumento de evaluación (sus propiedades psicométricas: fiabilidad y validez); el director del trabajo (perteneciente a la Complutense: evaluador interno) y la empresa encargada de la recogida y el análisis de datos (evaluador externo). Con la estrecha colaboración entre el evaluador interno y el externo se pretende paliar, al menos en parte, algunos de los inconvenientes de cada uno de ellos (posible falta de objetividad en el primero y falta de conocimiento en el segundo, entre otros), que han sido sacados a la luz por los distintos expertos en evaluación.
- Unos *datos descriptivos básicos* (media, mediana, moda, desviación típica, distribución de frecuencias y representación gráfica de las medias), referidos a los 29 elementos del CEPUC, junto con la manera adecuada de interpretarlos.
- La evaluación sintetizada en función de los *núcleos teóricos* a los cuales ya se hizo referencia.
- Una *evaluación comparada* (tanto en cifras como en su representación gráfica) *del profesor con el resto de unidades de análisis significativas* -en cuanto a docencia se refiere-: asignatura (si es compartida), departamento, centro, tipo de centro (experimental, no experimental, escuela universitaria) y universidad.
- Una *evaluación comparada del profesor consigo mismo a lo largo del tiempo*. De esta forma el profesor se puede convertir en *investigador* con respecto a su propia calidad docente, pues año tras año puede ir comprobando hasta qué punto los cambios que va introduciendo van dando un tipo de fruto determinado.

También en estas hojas introductorias aparecen dos matizaciones que juzgamos de relevancia. La primera hace referencia a la interpretación estadística. Supongamos que dos profesores obtienen en un determinado ítem o en un núcleo teórico una evaluación media de 4. Aparentemente estaríamos ante dos profesores iguales en cuanto a calidad docente se refiere. Sin embargo, si nos detenemos en, por ejemplo, la desviación típica o la moda, vemos que pueden ser muy diferentes. De hecho, en un caso, prácticamente todos los alumnos coinciden en otorgarle un 4, mientras que en el otro se produce una división extrema de la clase, hasta el punto de que la mitad lo valora con un 7 y la otra con un 1. Al tener en cuenta no sólo la media –semejante en ambos casos- sino también los otros índices, más concretamente el que hace referencia a la variabilidad –la desviación típica-, claramente discrepante, el juicio final es ahora muy diferente al inicial. Ya no hablaríamos

de dos calidades docentes muy semejantes sino de dos tipos de profesores considerablemente distintos, pese a su semejanza en las medias obtenidas.

La otra matización se refiere a la interpretación global de los resultados obtenidos. Se indica al final de la tercera hoja que este procedimiento de evaluación de la docencia es sólo uno de los múltiples posibles, aunque, como hemos tenido la oportunidad de comprobar, el más extendido dentro del mundo universitario, a escala internacional. De ahí que lo que el profesor obtiene no sea ni más, pero tampoco menos, que el conocimiento de las opiniones de los alumnos sobre la enseñanza que reciben.

Al final de estas hojas introductorias se le indica que el objetivo esencial de la evaluación es claramente formativo, es decir, lo que se pretende es ofrecerle una oportunidad, mediante el envío de los resultados de la evaluación, de optimizar su enseñanza, bien sea gracias a sus propios esfuerzos investigadores o bien a la ayuda del equipo de expertos que ha diseñado y dirigido la propia evaluación. Este equipo cuenta en su haber con todo tipo de evaluaciones (más de 4.000 profesores evaluados por curso académico): las excelentes, las buenas, las menos buenas, las medianas y las malas. Tomando como ejemplo y punto de partida las excelentes se puede tratar de que el resto se vaya acercando paulatinamente a esos niveles de excelencia, mediante la corrección de sus puntos más débiles (gracias a las múltiples técnicas hoy disponibles: aprendizaje por observación, grabación y corrección, adopción de perspectivas, entre otras), siempre y cuando previamente se haya establecido una cultura de la evaluación.

2.4. Los criterios consolidados de evaluación: el de la escala de valoración, el relativo espacial y el relativo temporal

Como acabamos de ver, a diferencia de otros sistemas de evaluación que suelen utilizar un único criterio de valoración, en éste cada profesor es evaluado por los alumnos en función *de tres claros criterios consolidados*. Gracias al primero, derivado de la propia escala tipo Likert de valoración, los profesores pueden detectar sus **puntos más fuertes y débiles**. Como se especifica en la primera hoja de la devolución de la información, para cualquiera de los índices de tendencia central (media, mediana o moda) se ha de seguir siempre un *principio general*: “*las puntuaciones por encima de 4 (punto medio de la escala valorativa) ponen de manifiesto una consideración positiva, mientras que las puntuaciones inferiores a ésta han de interpretarse en la dirección de una evaluación más bien negativa*”. El perfil docente de cada profesor que aparecerá recogido en la página 5 del modelo de devolución de la información (como ahí se indica, no corresponde a puntuaciones de ningún profesor real en concreto: el compromiso ético lo impide –era uno de los criterios de la metaevaluación-) nos permite detectar, de una forma muy gráfica, todos y cada uno de los puntos fuertes y débiles, siempre, como es lógico, dentro de los límites establecidos por los 29 ítems que componen el CEPUC intermedio. Esto mismo vale para los resultados que aparecen en la página 6. Aquí, todavía con más claridad, uno puede constatar de forma inmediata esos núcleos teóricos mejor y peor evaluados.

Ahora bien, imaginémosnos que un profesor obtuviese, según este criterio, una evaluación positiva, pues en todos los elementos y los núcleos ha conseguido un 5, dentro siempre de la escala de valoración que va de 1 a 7. Tanto la institución como él mismo, al menos en un primer momento, debieran de estar contentos. Sin embargo, con la utilización del *segundo criterio –el de la comparación de su evaluación con el resto*

de unidades de análisis significativas-, ese primer juicio puede cambiar, puede ser mucho más matizado. Siguiendo con el supuesto ejemplificador, imaginemos que el resto de unidades de análisis –el resto de compañeros de la asignatura, del departamento o del centro- hubieran obtenido una media empírica de 6 en todos los ítems y núcleos. Teniendo todo esto en cuenta, ya ese supuesto profesor no nos parecería en modo alguno tan bueno. Es más, si la aspiración de la institución fuera la de obtener una *calidad total*, muy probablemente dicho profesor tuviera que mejorar considerablemente su docencia o correría el riesgo de ver cancelado su contrato. Gracias, pues, a este segundo criterio, el primero puede ser considerablemente matizado y, por consiguiente, enriquecido.

Pero todavía nos queda un *tercer criterio*, que a su vez puede matizar y enriquecer a los dos primeros. Nos referimos al *comparativo en función del tiempo*. De nuevo, por razón de ilustración pedagógica, pongámonos en la situación de dos casos extremos: a) el de un profesor que habiendo comenzado con evaluaciones en las que predominaban los puntos débiles va transformándolos poco a poco hasta convertirlos en auténticos puntos fuertes –pasa de una evaluación, por ejemplo, en torno al 3 a otra en torno al 6-; b) el de un profesor cuyos puntos débiles van en aumento –de una evaluación cercana al 6 va camino de otra cercana al 3- a medida que van pasando los cursos académicos. Nosotros tenemos acceso imaginario a sus evaluaciones cuando ambas rozan el 5. Sin este tercer criterio volveríamos a cometer un craso error, si pensáramos que estamos ante dos profesores cuya calidad docente es semejante, pues ambos presentan evaluaciones en torno al 5. De hecho, con la ayuda de este tercer criterio, podemos afirmar que nuestro juicio sería muy diferente ante una supuesta institución académica que quisiera saber nuestra opinión de expertos en evaluación de la calidad docente. Aconsejaríamos sin ningún género de dudas, si todas las demás variables fueran iguales, una apuesta por el primero y un rechazo del segundo.

Para cualquier profesor, el poder contar con información referida a los tres criterios (el de la escala de valoración, el comparativo espacial y el comparativo temporal) le supone a todas luces una ayuda muy superior a la proveniente de cualquiera de ellos por separado. A partir de esta información, si lo desea puede pedir ayuda a los compañeros que destaquen precisamente en aquellos elementos en donde él es más débil (puede ir a observarlos a su aula, puede pedir permiso para grabarlos, puede entrevistarlos para que le expliquen las estrategias por ellos utilizadas, etcétera) y también puede por sí mismo elaborar hipótesis en torno a las posibles causas de sus puntos débiles y, con o sin ayuda experta, tratar de verificarlas mediante los cambios que pueda ir introduciendo en su docencia en cursos sucesivos. Las nuevas devoluciones de la información cobrarán así un nuevo significado, además del específico evaluativo que ya tenían, al convertirse en la prueba empírica para la aceptación o rechazo de sus propias hipótesis (*investigación en la acción*).

A todo lo anterior, creímos oportuno incorporar un *cuarto criterio*, centrado en la *autoevaluación* (ver apartado de Anexos: Anexo 5). Cada profesor cumplimenta 27 de los 29 ítems del CEPUC, cambiados únicamente en su redacción (*AUPRO*), a fin de que pueda constatar su concordancia o discordancia con las opiniones de sus alumnos. La conclusión general a la que es posible llegar es que ambos perfiles –el obtenido gracias a las opiniones de los alumnos y el conseguido con las opiniones del propio profesor- son semejantes (coinciden los puntos fuertes y los débiles), con el único matiz

de que, en general, las puntuaciones que se otorga cada profesor son *superiores* a las que le son dadas por sus alumnos.

3. Las opiniones de los profesores

Hasta aquí, con mejor o peor fortuna, he seguido las pautas ya marcadas por otros investigadores, sobre todo estadounidenses, a la hora de acercarme a la evaluación de la calidad de la enseñanza universitaria, siendo este enfoque extrapolable –*mutatis mutandis*–, a las enseñanzas no universitarias. De hecho, algunos de mis alumnos del curso de doctorado –profesores ellos a su vez - han utilizado el CEPUC para su propia evaluación o la de sus compañeros de centro, sin que el instrumento de evaluación variase significativamente en ninguna de sus propiedades formales. Ahora bien, desde un principio, siempre me pareció que este enfoque sólo nos proporcionaba información fidedigna de una parte –ciertamente esencial: la referida al alumnado–, sin que tuviésemos datos sobre el otro pilar básico –el del profesorado- del primer triángulo que vimos en el capítulo primero (alumnos, profesores y currículos: Figura 1). Por ello –por un mínimo de coherencia lógica y académica- no tuve más remedio que tratar de dar solución a este problema, mediante la elaboración de un nuevo instrumento de evaluación de la calidad a cumplimentar por los profesores.

3.1. La elaboración de un nuevo instrumento de evaluación y sus núcleos teóricos

Mientras que con respecto al primer cuestionario había, como se ha puesto de manifiesto anteriormente, abundantes fuentes de inspiración y, por supuesto, múltiples opciones para elegir si éste hubiese sido mi propósito, no ocurría lo mismo con el segundo cuestionario: apenas fue posible encontrar instrumentos bien elaborados y publicados. Ahora bien, dado que se pretendía dar a luz un modelo de evaluación circular, no quedaba más remedio que tratar de construir un segundo cuestionario, complementario del primero. También en esta situación, al igual que ocurrió con el CEPUC, se elaboró una primera versión amplia de 69 ítems que, sometidos a los correspondientes análisis factoriales exploratorios, nos proporcionó 6 factores que daban cuenta del 75% de la varianza. Por lo que respecta a su fiabilidad, el coeficiente theta de Carmines, alcanzó el valor de 0.98 para todo el cuestionario (*CECA*, en castellano, y *ASEQ*, en la versión inglesa). Estos primeros resultados eran claramente satisfactorios. No obstante, parecía oportuno contar, igualmente, con una versión reducida. Para ello, seleccionamos aquellos 33 ítems que formaban parte de los tres primeros factores, al ser éstos los que para nosotros y para la institución –la Universidad Complutense- gozaban de una mayor significación práctica y académica. Por lo que respecta a su fiabilidad, consistencia interna, el coeficiente alpha de Cronbach alcanzó un valor de 0.90 y el coeficiente theta de Carmines de 0.97. En cuanto a la estructura factorial, cabe indicar que los tres factores previstos *a priori* (*Satisfacción con las instituciones universitarias*; *Clima social intradepartamental*; *Relaciones con los estudiantes*) explicaban un 78 % de la varianza. Aunque un análisis minucioso aconsejaba el pulimento del *CECA-R* de 33 ítems (ver Fernández y Mateo, 1993, y Mateo y Fernández, 1995, para más detalles técnicos: análisis factoriales exploratorios y confirmatorios), la valoración general es positiva, pues confirma y apoya los resultados obtenidos en el primer estudio con la versión más amplia.

Como ya hemos visto con respecto al CEPUC, algo similar ha sucedido con el CECA a la hora de su utilización en la realidad. La versión intermedia (de 59 ítems: ver el apartado de Anexos: Anexo 3) es la más usada en la práctica. Además, para el sistema de devolución de la información hemos recurrido una vez más a los *núcleos teóricos*, que tienen el correspondiente respaldo estadístico (ver este modelo de devolución de la información en Anexos: Anexo 4.2). La explicación de este modelo de devolución de la información se hace innecesaria al ser semejante a la ya comentada para el CEPUC. Ahora bien, dado que en esta ocasión los resultados obtenidos son de dominio público, puesto que fueron recogidos en su día incluso por alguno de los principales periódicos españoles de ámbito nacional, se ofrece a continuación una síntesis de los mismos con el fin de que el lector, una vez más, pueda tomar decisiones acertadas sobre el modelo de evaluación circular que se le está describiendo, a la par que recordar y asimilar, ahora en la práctica, algunos de los conceptos básicos que se desarrollaron en la primera parte –la conceptual- y que son esenciales para cualquier modelo de evaluación que uno decida elegir (para una información más detallada, tanto de tipo metodológico como conceptual, puede consultarse Fernández, 1997; Fernández y Mateo, 1993, 1994; Fernández, Mateo y Muñiz, 1995; Mateo y Fernández, 1995).

Estos resultados son el fruto de tres investigaciones distintas, realizadas con un intervalo de tres cursos académicos entre ellas (1986/87, 1989/90 y 1992/93) y que tienen como común denominador la utilización del CECA en su versión intermedia (59 ítems). El primer estudio fue financiado por el Consejo de Universidades y nos permitió aplicar este cuestionario en una muestra representativa de las universidades españolas (800 profesores). En el segundo, la financiación corrió a cargo de la Universidad Complutense. Participaron un total de 2.547 profesores. En el tercero, financiado por la Universidad de Oviedo, participaron 655 profesores.

En la Tabla 4 se recogen los datos referidos al *primer núcleo teórico: Satisfacción ante las instituciones universitarias*.

Tabla 4

Valoración de las condiciones en que ejerce sus funciones docentes e investigadoras el profesorado

	MEDIAS Escala de 1 a 7		
	86 /87 CU*	89 /90 UCM	92 /93 UO
Satisfacción ante las instituciones universitarias			
1. Las condiciones materiales en que realizo mi trabajo son satisfactorias	3.0	3.2	3.9
2. Se me ofrecen posibilidades financieras para realizar mi investigación	2.4	2.6	3.2
3. Se me dan facilidades institucionales para realizar mis trabajos	2.8	2.8	3.3
4. Considero justamente remuneradas mis actividades como profesor	2.4	2.8	3.4
11. Existen criterios claros para evaluar las actividades investigadoras	2.2	2.4	2.5
12. Concordancia entre expectativas y realidad de ser profesor	2.8	3.1	3.3
13. La sociedad aprecia la labor realizada por el profesor universitario	2.6	2.8	3.1
14. Las instituciones universitarias estimulan mi perfeccionamiento como docente	2.4	2.5	2.5
16. Preparación institucional para ejercer mis funciones de investigador	2.5	2.7	2.8
17. Existen perspectivas favorables para realizar mi trabajo como docente	2.9	3.2	3.4
18. Las perspectivas futuras como investigador universitario son favorables	2.7	3.1	3.2
19. Facilidades institucionales para la resolución de problemas como profesor	2.8	2.9	3.1
20. Cuento con el tiempo adecuado para realizar mis tareas de investigación	3.1	3.2	3.3
22. Las instituciones universitarias fomentan mi actividad investigadora	2.4	2.8	2.8
34. Satisfacción general ante las condiciones institucionales		3.0	3.6

* CU = Investigación realizada en un tercio de universidades españolas para el Consejo de Universidades (hoy Consejo de Coordinación Universitaria).

UCM = Universidad Complutense de Madrid.

UO = Universidad de Oviedo.

Es importante señalar que hay *varias lecturas* posibles a la hora de interpretar estos datos. Comenzando por una de *carácter global*, cabe afirmar que los datos nos ofrecen una evaluación no excesivamente positiva de todo este núcleo por parte de los profesores, en los tres estudios. En ningún caso, la valoración dada a los 15 elementos sobrepasa el punto de corte de la positividad -4-, que es la media teórica de la escala de valoración, que va de 1 a 7 (1 indica siempre el grado más bajo del contenido de cada elemento y 7 el grado más alto). Como resulta lógico, éste es un factor condicionante importante a la hora de reinterpretar las propias evaluaciones realizados por los alumnos con respecto a la calidad docente de sus profesores. No es lo mismo que los profesores estén contentos que descontentos con las condiciones en que han de ejercer sus funciones docentes. Al menos, como hipótesis, parecería más lógico que las autoridades académicas exigiesen una mejor evaluación de la docencia a aquellos profesores que gozan de un contexto más favorable que a los que han de trabajar bajo mínimos –las medias empíricas no superan la media teórica de la escala de valoración-, como se constata gracias a los datos arriba expuestos. De hecho, ésta es una de las críticas (de ahí la evaluación muy negativa para las autoridades ministeriales por parte del profesorado, como veremos más adelante) realizadas contra el sistema de evaluación de los alumnos, pues se argumenta que se pretende realizar una evaluación como las de las mejores universidades estadounidenses pero en unas condiciones de trabajo (los españolas) que no tienen mucho que ver con las existentes al otro lado del Atlántico.

Esta primera visión general se puede completar con *otras lecturas más específicas*, a fin de perfeccionarla y enriquecerla. Traigamos a colación, sólo a título de ejemplo ilustrativo, alguna de ellas. Además de lo dicho, se puede constatar:

a) La baja y relativamente uniforme valoración de ciertos ítems

- * El 11: Existen criterios claros para evaluar las actividades investigadoras -2.2, 2.4, 2.5-.
- * El 14: Las instituciones universitarias estimulan mi perfeccionamiento como docente -2.4, 2.5, 2.5-,
- * El 16: Preparación institucional para ejercer mis funciones de investigador -2.5, 2.7, 2.8-.

Estos datos parecen reclamar, de forma clara y contundente -claro consenso de evaluación negativa entre los profesores de las distintas universidades-, una intervención ministerial para mejorar el contexto universitario español, dado que, por lo que respecta a la docencia, las instituciones universitarias no parece que se preocupen de estimular su calidad o excelencia.

b) La baja aunque dispar (entre la última evaluación y las dos anteriores) valoración de otros ítems

- * El 1: Las condiciones materiales en que realizo mi trabajo son satisfactorias -3.0, 3.2, 3.9-.
- * El 2: Se me ofrecen posibilidades financieras para realizar mi investigación -2.4, 2.6, 3.2-.
- * El 4: Considero justamente remuneradas mis actividades como profesor -2.4, 2.8, 3.4-.

A la luz de estos datos, vemos que las autoridades ministeriales tendrían aquí algunos *indicios* para saber hacia dónde debería dirigirse su *intervención* (hacia el contexto de la Universidad de Oviedo, pero mejorado), pues la situación actual, en general, todavía es negativa.

En la Tabla 5 se muestran los datos referidos a los núcleos teóricos del *Clima social intradepartamental y de las Relaciones con los alumnos*.

Tabla 5
Valoración del clima entre los miembros del departamento y de las relaciones con los alumnos

	MEDIAS		
	Escala de 1 a 7		
Clima social intradepartamental	86 /87	89 /90	92 /93
5.Me siento respaldado por los compañeros en las actividades como profesor	4.7	4.5	4.3
6. El contexto académico fomenta mi labor profesional	3.5*	3.3	3.4
7. Existe adecuada comunicación académica entre los miembros de mi departamento	3.8	3.7	3.1
8. Las relaciones con los compañeros de mi dpto. favorecen mi actividad académica	4.1*	4.2	3.7
26. Colaboración intradepartamental en relación a programas de investigación	3.9	3.7	3.3
33. Me siento apoyado en mis investigaciones por los compañeros del departamento	4.1	3.8	3.6
35. Clima intradepartamental	3.5	3.4	3.3
Relaciones con los alumnos			
21. Los alumnos manifiestan interés por la materia que imparto	4.4	4.6	4.4
24. Los alumnos consultan sus dudas en el horario destinado a tal efecto	3.8	3.7	3.2
25. Se tienen en cuenta las opiniones de los alumnos sobre la docencia	4.6	4.8	4.7
27. Considero las opiniones de los alumnos para establecer la metodología docente	4.7	5.2	4.9
28. Acomodo mi docencia a las peculiaridades de cada grupo de alumnos	4.9	5.3	5.1
29. El trabajo de los alumnos se ajusta a las demandas de mi asignatura	4.2	4.5	4.4
32. Valoración diferencial del alumnado de la calidad docente	4.2	4.9	4.6
36. Participación activa de los alumnos	3.6	3.3	2.9

* = Ligera modificación en la redacción del elemento en este estudio.

Por lo que respecta al *clima social* cabe constatar que si bien la valoración general es algo más positiva que la que reflejaba el primer núcleo, sin embargo, los aspectos evaluados negativamente predominan sobre los que obtienen una evaluación satisfactoria. Sólo un elemento (el 5: Me siento respaldado por los compañeros en las actividades que realizo como profesor) aparece con una valoración por encima del punto de corte -media teórica- en los tres estudios. Otros dos elementos (el 8: Las relaciones con los compañeros de mi departamento favorecen mi actividad académica, y el 33: Me siento apoyado en mis investigaciones por los compañeros del departamento) presentan valores en torno a la media teórica de la escala valorativa, cuando se tienen en cuenta conjuntamente los resultados de las tres investigaciones. Para el resto de elementos la evaluación es más bien negativa en los tres casos. Este primer juicio –lectura de carácter global- queda enriquecido –otras lecturas más específicas- cuando se tiene presente que el clima social parece seguir una línea descendente hacia la negatividad a medida que transcurre el tiempo. De nuevo aquí, hemos de recurrir a la lógica para inferir sin excesivos riesgos de equivocarnos que un clima social no excesivamente bueno entre los miembros de un departamento tal vez no sea el mejor contexto para desarrollar una excelencia docente. Además, se nos indica que en función de determinados factores, que se pueden controlar, ese clima puede ir en una u otra dirección.

En cuanto a las *relaciones con los alumnos*, los datos parecen avalar una evaluación global positiva –lectura global-, tanto desde la perspectiva comparativa con respecto a los dos núcleos significativos anteriores, como desde el análisis en sí mismo, dado que en la mayoría de los elementos las medias empíricas superan el promedio teórico de la escala valorativa en los tres trabajos. No obstante, si añadimos otras lecturas más específicas, nos encontramos con dos elementos (el 24: Los alumnos consultan sus dudas en el horario destinado a tal efecto, y el 36: Participación activa de los alumnos) cuyos valores están por debajo del punto de corte. Estos datos nos están señalando un tipo de sistema educativo “clásico” (de lección magistral), en donde el alumno recoge sus apuntes para ulteriormente examinarse de ellos. No parece que los profesores se quejen del comportamiento de sus alumnos, dada su pasividad, pero estos resultados nos debieran hacer pensar –a los responsables académicos, sobre todo- en si el sistema complementario de, por un lado, la función activa del profesor y, por el otro, la función pasiva de los alumnos, es el sistema más deseable para los contextos actuales de aprendizaje/enseñanza. La interacción se ha reducido al mínimo y el aprendizaje compartido brilla por su ausencia.

En la Tabla 6 se recogen los datos referidos tanto al núcleo teórico de *Selección y promoción del profesorado como al del bloque de elementos ponderativos*.

Tabla 6

Valoración de los sistemas de selección y promoción del profesorado, y de un conjunto de ítems ponderativos

	MEDIAS Escala de 1 a 7		
Selección y promoción del profesorado	86 /87	89 /90	92 /93
9. Los sistemas de selección del profesorado son los adecuados	2.6	2.7	2.7
10. Los sistemas de control de la actividad docente son correctos	2.3*	2.8	2.8
15. Los sistemas de promoción del profesorado son correctos	2.4	2.4	2.4
23. El sistema de funcionariado es el correcto para ser profesor	3.0	3.1	3.5
30. Los contratos laborales posibilitarían mejorar las funciones del profesor	4.2	3.9	3.7
31. Conveniencia de un sistema "objetivo" para la valoración de la investigación	5.4	5.2	5.5
37. Los sistemas actuales de selección y promoción son adecuados		2.5	2.6
Bloque de elementos ponderativos			
38. Los alumnos se cuestionan la utilidad de los contenidos impartidos	4.1	4.2	4.0
39. La evaluación docente de los alumnos es útil a la institución	4.0	4.2	3.5
40. El número de alumnos es el correcto para prepararlos profesionalmente	2.1	2.6	2.8
41. Los alumnos proponen formas alternativas al modo de dar las clases	2.3	2.4	1.8
42. Existe colaboración intradepartamental en la programación docente	2.7	3.0	2.6
43. Enseño la asignatura para la que realmente me encuentro más capacitado	5.5	6.0	5.6
44. Puedo enfocar la asignatura desde la perspectiva que juzgo más conveniente	5.3	6.0	5.8
45. La formación recibida para ejercer como profesor es suficiente	5.1*	5.0	4.5

*= Ligera modificación en la redacción del elemento en este estudio.

Ante todo se constata –lectura global- que los sistemas de selección y promoción actuales del profesorado no satisfacen en absoluto a la inmensa mayoría de los profesores de la universidad española. Esto es a todas luces grave. Ahora bien, gracias a las lecturas más concretas, vemos que se vislumbra una solución: la introducción de sistemas "objetivos" de evaluación. Sin estos sistemas objetivos de evaluación, lo lógico es que la picaresca haga su aparición y el clima entre los compañeros docentes quede enrarecido. Incluso aparecen más matices: se podría pensar en los contratos laborales como una alternativa (los valores de las medias son algo más elevados) que podría convivir con el sistema de oposiciones a funcionario que es el sistema que ha prevalecido, como *desideratum*, en la universidad española hasta nuestros días.

Cuando nos detenemos en el núcleo teórico de los *elementos ponderativos*, nos encontramos una vez más con la ya bien constatada semejanza del perfil valorativo en los tres estudios. Los ítems bien evaluados son tan concordantes como los menos bien evaluados. Así, los tres ítems últimos presentan unos valores tremendamente altos, pues se refieren a la visión que los propios profesores tienen de sí mismos: [el 43 (Enseño la asignatura para la que realmente me encuentro más capacitado -5.5, 6.0, 5.6-); el 44 (Puedo enfocar la asignatura desde la perspectiva que juzgo más conveniente -5.3, 6.0, 5.8-); el 45 (La formación recibida para ejercer como profesor es suficiente -5.1, 5.0, 4.5-)], mientras que otros tres ítems, referidos a los alumnos o a la colaboración entre pares, muestran unos valores más bien bajos: [el 40 (El número de alumnos es el correcto para prepararlos profesionalmente -2.1, 2.6, 2.8-); el 41 (Los alumnos proponen formas alternativas al modo de dar las clases -2.3, 2.4, 1.8-); el 42 (Existe colaboración intradepartamental en la programación docente -2.7, 3.0, 2.6-)]. En el primer caso nos volvemos a encontrar con ese sesgo que suele ser típico del profesorado cuando se ha de autoevaluar: sobreestimación. Ya aparecía cuando se le aplicaba el CEPUC especialmente diseñado para que pudiera ser respondido por los profesores. Como se ve, con uno u otro método, desde una u otra perspectiva, cuando la fuente de información sobre su comportamiento es el propio profesor, los resultados se repiten (replicación técnica). En el segundo caso, aparece un doble matiz. Por un lado, no hay participación por parte del alumnado a la par que el número de éstos por clases es excesivo (¿están relacionadas ambas variables? ¿Una es causa de la otra?). Por otro, se ratifica (se replica) el resultado que ya conocíamos: el clima entre el profesorado de un mismo departamento no es el adecuado cuando uno piensa en términos de calidad o excelencia docente.

Especial mención merecen los ítems primero (el 38: Los alumnos se cuestionan la utilidad de los contenidos impartidos -4.1, 4.2, 4.0-) y segundo (el 39: La evaluación docente de los alumnos es útil a la institución -4.0, 4.2, 3.5-) de este bloque. En el primer caso, se percibe una clara insatisfacción con lo dado en las clases. Tal vez alguien debiera preguntarse el porqué. En el segundo, parece que no todos los sistemas de evaluación son iguales. Esto es especialmente relevante para los planteamientos defendidos en este libro. Si bien, se ha de dar la bienvenida a la pluralidad de sistemas evaluadores, sin embargo esto no debiera significar que todos son igualmente buenos, eficaces o eficientes. Así, mientras en un caso la media empírica supera la media teórica, en otro está por debajo, mientras en un tercero se da justamente una total coincidencia.

En la Tabla 7 se muestran los datos referidos al núcleo teórico del *Funcionamiento de los servicios institucionales*.

Tabla 7
 Valoración de los distintos servicios institucionales

	MEDIAS		
	Escala de 1 a 7		
Funcionamiento de los servicios institucionales	86 /87	89 /90	92 /93
46. De la biblioteca	4.3	4.6	4.2
47. De los laboratorios	3.6	3.7	4.0
48. De la hemeroteca	3.9	4.2	3.7
49. De los medios informáticos	3.8	3.5	4.2
50. De los medios audiovisuales	3.1	3.3	3.8
51. De la secretaría	4.8	4.6	5.0
52. Condiciones físicas del inmueble	3.3	3.3	4.1
53. Actividades culturales organizadas por la institución	3.1	3.3	3.1
54. Del personal auxiliar	3.8	3.9	3.8
55. De los decanatos	4.7	4.7	4.2
56. Del rectorado	3.6	4.4	4.1
57. De las instalaciones deportivas	3.3	3.8	3.7
58. De los comedores y cafeterías	3.4	3.1	3.8
59. Del Ministerio de Educación y Ciencia	*	3.0	2.9

* No se incluyó tal valoración en este estudio.

Una vez más es importante reseñar la semejanza del perfil valorativo de las tres investigaciones, con un intervalo entre ellas de tres cursos académicos y con responsables de la recogida de la información muy diferentes: dos empresas distintas en los dos primeros estudios y la propia Universidad de Oviedo en el tercero. Difícil, pues, imaginar que nos encontremos ante unos resultados sesgados. Más bien, por el contrario, parecen constituirse en una fiable y válida radiografía de las condiciones en que han de desarrollar sus funciones los docentes universitarios españoles. La sencillez en la presentación de los datos hace innecesarios comentarios detallados con respecto a este último núcleo. Bastará, por tanto, con señalar, a título de pauta ilustrativa, los valores que se encuentran en los dos extremos. En el polo más positivo, en los tres estudios, aparecen las secretarías y en el más negativo el Ministerio de Educación y Ciencia (MEC), en los dos estudios en los que se incluyó este elemento.

Sin duda, los resultados referidos al último ítem merecen una consideración especial, pues el MEC es a todas luces “suspendido” por los profesores. Además, algunos ítems del cuestionario de evaluación nos ofrecen pistas sobre las razones del suspenso. Son las autoridades del MEC, por ejemplo, las responsables de los sistemas de selección y promoción del profesorado y, como vimos, los valores que alcanzaban aquellos ítems eran bastante semejantes a los que ahora presenta este último ítem. Por si esto fuera poco, cuando desglosamos los resultados globales de estos ítems en función de los distintos centros de la Complutense, obtenemos un perfil (una radiografía: véase Tabla 8) sin manifiestos dientes de sierra, es decir, bastante plano, lo que nos indica que se da una coincidencia prácticamente total (suspenso rotundo: ninguna media empírica llega a 3) sin importar el tipo de centro (experimentales, no-experimentales o escuelas universitarias) o las ideologías de los profesores.

Tabla 8

Valoración del ítem 37, desglosada en función de los distintos centros académicos de la Universidad Complutense

Ítem 37: Los sistemas actuales de selección y promoción del profesorado son los adecuados			
Unidades evaluadoras	Media	Moda	Des. Típ.
Universidad Complutense (todos los centros)	2.5	1	1.4
Centros experimentales	2.4	1	1.3
Centros no experimentales	2.5	1	1.4
Escuelas universitarias	2.7	2	1.4
Facultad de Filosofía y Ciencias de la Educación	2.6	1	1.3
Facultad de Psicología	2.8	2	1.4
Facultad de Filología	2.3	1	1.2
Facultad de Geografía e Historia	2.4	1	1.4
Facultad de Ciencias Químicas	2.3	1	1.3
Facultad de Ciencias Físicas	2.6	2	1.3
Facultad de Ciencias Matemáticas	2.7	1	1.6
Facultad de Ciencias Biológicas	2.4	1	1.5
Facultad de Ciencias Geológicas	2.2	1	1.2
Facultad de Medicina y Hospitales	2.3	2	1.1
Facultad de Farmacia	2.3	1	1.3
Facultad de Veterinaria	2.6	2	1.5
Facultad de Odontología	2.4	2	1.3

Facultad de Derecho	2.6	*	1.3
Facultad de Ciencias Políticas y Sociología	2.5	1	1.3
Facultad de Ciencias Económicas y Empresariales	2.7	2	1.3
Facultad de Ciencias de la Información	2.8	3	1.4
Escuela Universitaria de E.G.B. M ^a Díaz Jiménez	2.7	1	1.6
Escuela Universitaria de E.G.B. Pablo Montesino	2.5	2	1.2
Escuela Universitaria de Estudios Empresariales	2.9	2	1.3
Escuela Universitaria de Óptica	2.7	3	1.2
Escuela Universitaria de Estadística	2.6	3	1.1
Escuela Universitaria de Enfermería, Fisioterapia y Podología	2.4	1	1.8
Escuela Universitaria de Trabajo Social	2.4	2	1.2

* = No existe una única moda

Las implicaciones de estos datos saltan a la vista. ¿Por qué se sigue aplicando una normativa legal que, según la mayoría de los profesores –la moda en la escala de evaluación es muy baja: en la mitad es 1; en el resto un 2 ó un 3- es pernicioso para la universidad, siendo así que incluso dentro de los resultados totales se dan claras pistas de solución –los sistemas objetivos de selección y promoción- a la hora de encaminarse hacia la calidad o la excelencia? Otras muchas preguntas, que el atento lector puede formularse sin la más mínima dificultad, podrían ser dignas continuadoras de esta primera. Estos datos todavía cobran mayor significación al tener en cuenta que, cuando se realiza este análisis desglosado en función de los centros para todos y cada uno de los 56 ítems del cuestionario, los perfiles difieren considerablemente no sólo en cuanto a la evaluación –positiva, negativa, en torno a la media teórica-, sino también en cuanto al tipo: planos (como el que acabamos de ver) o con muchos dientes de sierra (véase la Tabla 9, como ejemplo).

Tabla 9

Valoración del ítem 3 de Servicios (Hemeroteca), desglosada en función de los distintos centros académicos de la Complutense

Hemeroteca			
Unidades evaluadoras	Media	Moda	Des. Típ.
Universidad Complutense (todos los centros)	4.2	6	2.0
Centros experimentales	4.6	6	1.9
Centros no experimentales	3.8	5	1.9
Escuelas universitarias	3.3	1	1.9
Facultad de Filosofía y Ciencias de la Educación	2.5	1	1.6
Facultad de Psicología	5.2	6	1.4
Facultad de Filología	2.7	1	1.9
Facultad de Geografía e Historia	3.1	1	1.8
Facultad de Ciencias Químicas	4.6	5	1.4
Facultad de Ciencias Físicas	5.0	6	1.5
Facultad de Ciencias Matemáticas	5.9	6	1.0
Facultad de Ciencias Biológicas	3.2	1	1.9
Facultad de Ciencias Geológicas	6.3	6	0.6
Facultad de Medicina y Hospitales	3.4	1	2.1
Facultad de Farmacia	3.1	1	1.8
Facultad de Veterinaria	5.2	6	1.6

Facultad de Odontología	5.2	6	1.5
Facultad de Derecho	3.6	5	1.9
Facultad de Ciencias Políticas y Sociología	3.1	1	1.8
Facultad de Ciencias Económicas y Empresariales	4.9	6	1.8
Facultad de Ciencias de la Información	4.3	5	1.6
Escuela Universitaria de E.G.B. M ^a Díaz Jiménez	3.6	*	1.7
Escuela Universitaria de E.G.B. Pablo Montesino	3.4	1	1.9
Escuela Universitaria de Estudios Empresariales	4.2	2	1.9
Escuela Universitaria de Óptica	2.7	*	1.2
Escuela Universitaria de Estadística	3.8	1	2.4
Escuela Universitaria de Enfermería, Fisioterapia y Podología	2.5	1	1.9
Escuela Universitaria de Trabajo Social	2.5	1	1.7

* = No existe una única moda

Es la consideración de estos dos aspectos conjuntamente –evaluación y tipo- lo que nos garantiza una mejor y mayor riqueza interpretativa. El tipo plano, debido a la concordancia de juicio de los evaluadores, nos garantiza la “objetividad” (es muy improbable que todos los evaluadores estén equivocados: Tabla 8), mientras que el perfil de picos de sierra nos permitiría diseñar programas de intervención con el objetivo de que lo negativamente evaluado, gracias al conocimiento de las variables que inciden en lo bien evaluado, pudiera ser mejorado (Tabla 9).

3.2. La circularidad

A la luz de estos datos, ya es posible llevar a cabo una valoración fundamentada tanto de las características específicas del modelo de evaluación circular (los alumnos evalúan al profesorado, éste a los alumnos y a las instituciones académicas, éstas a los profesores: en su docencia e investigación), como de las ventajas y desventajas de este modelo con respecto a los demás. Esta tarea, como es obvio, la dejamos en manos del lector. Para ello, como botón de muestra, destaco y resumo los elementos esenciales de la realidad de la circularidad evaluativa dentro del modelo expuesto.

Resulta obvio y lógico que el profesor tenga que evaluar el aprendizaje de sus alumnos. Siempre ha sido así y no parece que haya objeciones serias al respecto. Por eso desde este modelo no se insiste en ello. La única recomendación es que se haga de forma fiable y válida, lo cual, por desgracia, no siempre sucede (quedaría mucho por hacer a este respecto).

Menos generalizada (aunque se camina en esa dirección) está la evaluación llevada a cabo por el alumno con respecto a los comportamientos del profesor. Más allá de la reificación dimensional (Marsh puede ser un claro representante, con sus 9 y sólo 9 dimensiones), se han propuesto, desde nuestro modelo, unos núcleos teóricos -7 más uno formado por ítems de carácter general-, que de hecho se materializan en dos consistentes dimensiones y en dos claros conglomerados, más un ítem de carácter general. A la luz de los múltiples y diversos análisis estadísticos se puede afirmar que el instrumento de evaluación goza de buenas propiedades formales –altos índices de consistencia interna y dimensiones y conglomerados que se corresponden con los previstos *a priori*-, al menos para los ámbitos universitarios, aunque la extrapolación a la educación secundaria ha proporcionado igualmente buenos resultados: semejantes a los obtenidos en la universidad. Pese a ello, y a que estas dimensiones explican un alto porcentaje de varianza, somos

conscientes de que se puede mejorar el cuestionario y de que se pueden añadir elementos útiles para captar matices todavía no valorados por este instrumento.

Mucho menos frecuente, tanto a escala internacional como nacional, es la evaluación realizada por el profesorado. Éste puede y debe evaluar las condiciones laborales institucionales, el clima en el que se ha de desarrollar la docencia y la investigación, las relaciones con los alumnos (no solamente, pues, lo aprendido). Caben pocas dudas de que estas variables pueden incidir y, por consiguiente, condicionar la evaluación realizada por los alumnos, por lo que al considerarla en sí, sin este contexto, se nos privaría de una visión más extensa, más rigurosa, más exacta y, por tanto, más justa de la misma. A su vez, es muy difícil realizar una evaluación rigurosa y justa de la producción académica del profesorado –su calidad docente y su excelencia investigadora– sin tener en cuenta las condiciones contextuales en las que lleva a cabo dicha producción académica, aspecto éste que, lamentablemente, suele ser pasado por alto. Además, parece no ser el camino más aconsejable, cuando el puerto de llegada es la calidad y la excelencia, el que tiene por objetivo sacar a la luz leyes que no tienen en cuenta las opiniones de aquellos para los que se legisla. Los resultados que se han ofrecido son contundentes al respecto: suspenso para estas autoridades ministeriales.

Cuando se materializa, pues, la circularidad –cuando se cruzan los datos– el panorama que se nos presenta es de una riqueza de matices impresionante: las *radiografías* y los *perfiles* resultan ser más nítidos, fiables y válidos, tanto cuando aparece la concordancia de evaluaciones desde las distintas perspectivas –alumnos, profesores, autoridades académicas–, pues nos garantiza cierta “objetividad”, como cuando se manifiesta una clara discordancia (recuérdese el caso de la evaluación de las tutorías o el de la producción científica del profesorado), pues nos permite desvelar una realidad, hasta el momento desconocida (¿qué opinan los alumnos?, ¿qué opinan los profesores?, ¿cuáles son las razones de la discrepancia?), a partir de la cual será algo más fácil diseñar *programas de intervención*, siempre en aras de la mejora de lo evaluado, en la dirección de la calidad total.

4. La metaevaluación

Aunque pueda parecer, al menos a primera vista, que el modelo de evaluación circular goza de unas características superiores a las de los modelos que denominaré unidireccionales (evaluación del alumno por un lado y del profesor por otro), sin embargo, conviene contar con algunos datos empíricos metaevaluativos que avalen esa aparente superioridad lógica (Joint Committee on Standards for Educational Evaluation, 1994; Murray, 1984; Stufflebeam, 1978).

Ello nos obligó a la elaboración de un nuevo instrumento de valoración (*EPVA: Evaluación del Profesorado de la Valoración de los Alumnos*), compuesto de 7 ítems (véase para un análisis más detallado de las propiedades formales del mismo Fernández, Mateo y Muñiz, 1996). Baste decir aquí que goza de una razonable consistencia interna (theta de Carmines = 0.92) y de una satisfactoria validez de constructo: aparecieron las dimensiones previstas. Entre los tres factores que surgieron dan cuenta del 99.72% de la varianza empírica.

La aplicación del EPVA, tanto con profesores de la Universidad Complutense como con los de la Universidad de Oviedo, nos permite dar ahora cumplida cuenta de aquellas hipótesis que recibieron claro apoyo empírico:

- los *profesores asumen mejor la finalidad formativa de la evaluación que la sumativa*, aunque incluso ésta última es asumible y defendible;
- los *tipos de evaluación* llevados a cabo en cada contexto van a condicionar las valoraciones –*metaevaluaciones*- realizadas por los profesores, tanto sobre la finalidad formativa de la evaluación como sobre la sumativa.

Los datos empíricos que se presentan a continuación (Tablas 10 y 11) son ilustrativos de lo dicho (en primer lugar, los obtenidos en la Complutense –Tabla 10-y, en segundo lugar, los comparativos entre ambas universidades –Tabla 11-: véase todo ello de forma más extensa en Fernández, 1995; Fernández, Mateo y Muñiz, 1996).

Tabla 10

Metaevaluación: valoración por parte del profesorado de la evaluación realizada por los alumnos en la Universidad Complutense de Madrid

Ítems	Media	Moda	Des. Típica
1. Los resultados de la evaluación docente que se me entregan permiten detectar los puntos fuertes y débiles de mi docencia	4.7	5	1.6
2. Gracias a la comunicación de los resultados es posible modificar los aspectos menos positivos de mi enseñanza	4.8	5	1.6
3. La evaluación me ha sido útil para realizar modificaciones en mi manera de impartir las clases	4.6	5	1.6
4. Resulta conveniente un servicio de ayuda para cualquier profesor que quiera modificar su sistema de enseñanza	4.8	6	1.8
5. Estaría dispuesto a participar en actividades destinadas a la mejora de mi actividad docente	5.5	7	1.7
6. Estimo que se debieran utilizar los resultados como parte a tener en cuenta en los procesos de promoción del profesorado	4.4	7	2.1
7. Debiera utilizarse este tipo de evaluación como criterio fundamental para la concesión de quinquenios (evaluación de la calidad docente)	4.0	1	2.1

Como siempre, ante cualquier tipo de resultados varias lecturas son posibles y aconsejables. En este caso, comprobamos que los ítems referidos al aspecto formativo de la evaluación presentan una evaluación más positiva, si nos atenemos a la media empírica: en un sentido estricto de finalidad formativa, los tres primeros y en un sentido más amplio, los 5 primeros. Cuando pasamos al análisis de los dos últimos, de tipo sumativo, las medias bajan. Ahora bien, esta lectura es a todas luces insatisfactoria, pues una ojeada a las modas, nos indica que, al menos con respecto a los dos últimos ítems, se necesita llevar a cabo alguna precisión. ¿Por qué el 7 en el ítem 6? Porque los propios profesores, en varios ítems del CECA –de nuevo cruce de datos-, nos indicaron que no les satisfacían los sistemas de promoción del profesorado –eran muy subjetivos: la decisión final dependía más de quién estuviera formando parte de los tribunales evaluadores que de los méritos del

candidato-, y que deseaban un criterio más objetivo. Ahora ya sí se entiende el 7: reclamo de la “objetividad” frente a la “subjetividad”. ¿Por qué el 1 del ítem 7? Porque en la actualidad el quinquenio (un mínimo aumento salarial por cada lustro de docencia) se concede a todo el mundo por el mero hecho de solicitarlo. Buena parte del profesorado, como es lógico, no se quiere arriesgar a perderlo. Pese a ello, la media empírica de 4 nos indica que son muchos los profesores que, pese a la total gratuidad actual, prefieren que lo obtenido tenga su correspondiente justificación empírica.

¿Con esto habríamos acabado las lecturas? No. Si ahora nos fijamos en las desviaciones típicas, nos encontramos con la semejanza de valores para los primeros 5 ítems y también la igualdad en los dos últimos, aunque con claras diferencias entre ambos conglomerados. La variabilidad es superior en los ítems referidos al aspecto sumativo de la evaluación. Si todavía quisiéramos más precisión y riqueza, sólo nos quedaría recurrir a la distribución de frecuencias que, como ya hemos visto, siempre aparece en nuestro modelo de devolución de la información y que aquí hemos suprimido al no ser un informe remitido a los afectados o a las autoridades académicas, que son las personas capacitadas para poder elaborar programas de perfeccionamiento docente.

¿Sería ya ésta la última lectura posible? No. El desglose de estos resultados en función de determinadas variables demográficas de los participantes (sexo, rango, años de experiencia docente, categoría –funcionario frente a no funcionario-, prestigio, etcétera) nos permitiría desvelar muchos más matices todavía. En la realidad, contamos con esos datos y hemos emitido informes teniendo en cuenta algunos de estos desgloses: catedráticos y titulares de universidad; catedráticos y titulares de escuelas universitarias; numerarios y no numerarios (Fernández, 1995).

Cuando, además de lo dicho y hecho, nos atrevemos a realizar estudios comparativos, de nuevo los matices aumentan (véase Tabla 11).

Tabla 11

Comparación entre las medias empíricas de cada ítem del EPVA, obtenidas con profesores participantes de las universidades Complutense y de Oviedo

Ítems	Complutense	Oviedo	t	p	d
1	4.70	3.95	9.81	.0000	.44
2	4.80	4.13	8.57	.0000	.38
3	4.57	3.76	10.02	.0000	.45
4	4.80	4.70	1.07	.2831	.05
5	5.49	5.45	0.47	.6360	.02
6	4.40	3.89	5.30	.0000	.24
7	3.96	3.44	5.42	.0000	.24

Dejando de lado la posible interpretación de los datos de la Universidad de Oviedo, desde una perspectiva semejante a lo que acabamos de llevar a cabo con los resultados de la Complutense, nos vamos a centrar única y exclusivamente en los aspectos comparativos. Ellos nos manifiestan que:

- En los tres primeros ítems aparecen claras diferencias estadísticamente significativas, con un tamaño del efecto medio. Esto nos indica que los

profesores de la universidad de Oviedo juzgan que los resultados de las evaluaciones realizadas por los alumnos no les son tan útiles como a sus colegas de la Complutense, a la hora de detectar sus puntos fuertes y débiles y, por tanto, como ayuda para su perfeccionamiento como docentes. Muy probablemente sea el tipo de devolución de la información el que determine, al menos en parte, estas percepciones diferentes. Luego, si esto es cierto, *los diferentes tipos y modelos de evaluación sí parecen marcar diferencias significativas*.

- De hecho, la semejanza de resultados (no aparecen diferencias estadísticamente significativas, siendo los tamaños del efecto ínfimos) entre las dos universidades en los dos ítems siguientes, apoyaría, al menos de forma indirecta, la interpretación dada al párrafo anterior.
- Si no se está muy satisfecho con el tipo de evaluación y el formato de devolución de los resultados, cabría esperar en buena lógica que las valoraciones en los dos ítems siguientes fuesen bajas, puesto que de aspectos sumativos de la evaluación estamos hablando. Así ha sucedido, pues los valores promedio están por debajo del punto de corte teórico (el 4: media teórica de la escala de valoración). De nuevo aquí, las diferencias resultan estadísticamente significativas, aunque con unos tamaños de los efectos ligeramente inferiores a los encontrados para los tres primeros ítems. Recibiría así un nuevo apoyo empírico la hipótesis de que *el tipo y modelo de la evaluación puede condicionar la metaevaluación de los afectados*.

Estos resultados parecen ofrecer un considerable respaldo empírico a la postura ya expuesta en la primera parte –la conceptual- de que la cuestión importante y relevante no es la de un sí o un no a la evaluación, sino la referida al *tipo de evaluación*, es decir, a las buenas frente a las menos buenas o malas evaluaciones, sean éstas de tipo cualitativo o cuantitativo.

5. Algunas cuestiones en torno a posibles sesgos

Para que las evaluaciones sean creíbles (uno de los múltiples aspectos que se valoran a través de la metaevaluación), es necesario que el evaluador sea capaz de poder responder a todas y cada una de las cuestiones que los evaluados le puedan presentar. Las preguntas, desde este punto de vista, siempre son pertinentes –incluso cuando puedan ser malintencionadas o fruto de la ignorancia más supina-, mientras que la negativa a dar respuestas siempre resulta impertinente. Ahora bien, no todas las preguntas tienen igual valor: algunas son relevantes, desde un punto de vista de la evaluación, mientras que otras pueden resultar bastante irrelevantes. Aquí, me voy a centrar en algunas que destacan por su frecuencia: son bastantes los profesores e investigadores que quieren saber la mayor o menor incidencia de variables como el curso, el sexo, el turno, la asistencia a clase, la duración de la asignatura o el que los evaluadores sean o no repetidores, sobre la evaluación de la calidad de la enseñanza, etcétera (Beran y Violato, 2005; Centra y Gaubatz, 2000; Cohen, 2005; Marsh y Roche, 2000). Una información bastante completa sobre todas estas cuestiones puede consultarse en Fernández (1993).

Con respecto al *curso*, no es infrecuente encontrarse con profesores universitarios que afirman que es difícil que los alumnos de los primeros cursos puedan evaluar su calidad docente. Los datos recogidos en la Universidad Complutense, como se puede comprobar a través del análisis de la Tabla 12, desmienten esta hipótesis. Al menos dos lecturas son posibles de estos resultados. La primera, *en vertical*, nos indica que todos los cursos evalúan positivamente la enseñanza recibida en todos los ítems (valores iguales o superiores a 4, que es el punto de corte -la media teórica de la escala de valoración-, a la hora de dicotomizar una evaluación como positiva o como negativa). En una segunda lectura complementaria, la realizada en *horizontal*, comprobamos que los cinco perfiles son extremadamente semejantes. Los puntos más fuertes y los más débiles coinciden, así como los puntos intermedios, sabiendo que estamos siempre dentro de la zona de la valoración positiva, es decir, igualando o superando los valores empíricos de la media teórica. Dentro de los puntos fuertes nos encontramos con que en el ítem 2 todas las medias empíricas son 5 ó en el ítem 18 todas las medias son 5.4. Si nos vamos al otro extremo, siempre dentro del campo de la evaluación positiva, en el ítem 24 todas las medias son 4.2, excepto la de quinto curso que es 4.3 ó en el ítem 26 todas las medias empíricas son 4.1, excepto en primero que es 4.0. En todos los casos las desviaciones típicas son prácticamente iguales para cada ítem.

Tabla 12

La evaluación de la calidad docente de la Universidad Complutense en función de los distintos cursos

CURSO	1°	2°	3°	4°	5°
TOTAL EVALUACIONES	57.627	53.714	45.684	42.115	35.650
ÍTEMS (formato reducido)	Media	Media	Media	Media	Media
1. Su forma de enseñar facilita la comprensión	4.3	4.4	4.4	4.5	4.4
2. Muestra interés por las inquietudes del alumno	5.0	5.0	5.0	5.0	5.0
3. Motiva a preguntar sobre cualquier duda	4.5	4.6	4.6	4.6	4.5
4. Despierta interés por la asignatura	4.2	4.2	4.2	4.3	4.3
5. Se muestra accesible	4.9	5.0	5.0	5.0	5.0
6. Facilita la expresión de las ideas del alumno	4.6	4.6	4.7	4.7	4.7
7. Alude a las implicaciones de los enfoques	4.8	4.8	4.8	4.9	4.8
8. Establece interrelación entre los temas	4.8	4.8	4.7	4.8	4.7
9. Refleja la utilidad de los contenidos	4.6	4.6	4.6	4.7	4.6
10. Manifiesta entusiasmo por lo explicado	5.1	5.1	5.1	5.1	5.1
11. Ahora siento mayor interés por esta área	4.3	4.3	4.3	4.5	4.4
12. Responde con precisión	4.7	4.7	4.6	4.7	4.6
13. Mantuvo mi atención durante las clases	4.2	4.2	4.2	4.3	4.2
14. Explica de forma coherente	4.7	4.7	4.6	4.7	4.6
15. Atrae la atención del alumno sobre la materia	4.6	4.6	4.5	4.6	4.5
16. Explica con claridad	4.5	4.5	4.5	4.6	4.5
17. Estimula la participación	4.2	4.3	4.3	4.3	4.3
18. Se muestra competente en la materia	5.4	5.4	5.4	5.4	5.4
19. Adecuada preparación de la clase	5.5	5.4	5.3	5.4	5.3
20. Es constante en la asistencia a clase	6.1	6.0	6.0	6.0	6.0
21. Ajuste de lo explicado al programa	5.5	5.4	5.4	5.4	5.3
22. El examen permite reflejar lo aprendido	4.5	4.4	4.4	4.4	4.4

23. Examina sobre lo esencial	4.9	4.8	4.7	4.7	4.7
24. Las notas reflejan lo aprendido	4.2	4.2	4.2	4.2	4.3
25. La bibliografía sirve para preparar la materia	4.4	4.4	4.5	4.5	4.5
26. Me gustaría tenerlo de nuevo como profesor	4.0	4.1	4.1	4.2	4.1
27. Eficacia del profesor	4.8	4.8	4.7	4.7	4.7
28. Calificación comparativa con otros profesores	4.8	4.7	4.7	4.7	4.7
29. Calificación comparativa con otras asignaturas	4.5	4.6	4.5	4.7	4.5

Dada la importancia que últimamente está tomando la *realidad del género*, no es de extrañar que los profesores y profesoras quieran saber si la evaluación de la calidad docente puede estar significativamente condicionada por la variable sexo de los alumnos. Los datos que aparecen recogidos en la Tabla 13 nos permiten sacar una conclusión de carácter general: las mujeres suelen emitir un juicio valorativo ligeramente más favorable que los varones en 27 de los 29 ítems. Sólo en dos –el 26: Me gustaría tenerlo como profesor y el 28: Calificación comparativa con otros profesores–, los dos sexos coinciden. Cuando se realiza un análisis más exhaustivo (véase Fernández y Mateo, 1998), en el que se tiene en cuenta tanto la variable sexo de los alumnos como de los profesores y, en cuanto al análisis de datos, tanto las posibles diferencias estadísticamente significativas como, sobre todo, los tamaños de los efectos, la conclusión a la que se llega, como no podía ser de otro modo –recuérdense los datos obtenidos con el DPU: alumnos y profesores consideran esta variable irrelevante–, es que la variable sexo no condiciona significativamente la valoración de la calidad docente.

Tabla 13

La evaluación de la calidad docente de la Universidad Complutense en función de las variables sexo de los alumnos y turno

VARIABLES SEXO Y TURNO	Varones	Mujeres	Mañana	Tarde	Noche
TOTAL EVALUACIONES	80.464	155.331	159.299	66653	9.234
ÍTEMS (formato reducido)	Media	Media	Media	Media	Media
1. Su forma de enseñar facilita la comprensión	4.3	4.4	4.4	4.4	4.6
2. Muestra interés por las inquietudes del alumno	4.9	5.1	5.0	5.0	5.3
3. Motiva a preguntar sobre cualquier duda	4.4	4.6	4.5	4.5	4.8
4. Despierta interés por la asignatura	4.2	4.3	4.2	4.3	4.6
5. Se muestra accesible	4.9	5.0	4.9	5.0	5.4
6. Facilita la expresión de las ideas del alumno	4.5	4.7	4.6	4.7	5.1
7. Alude a las implicaciones de los enfoques	4.7	4.9	4.8	4.8	5.0
8. Establece interrelación entre los temas	4.7	4.8	4.8	4.8	5.0
9. Refleja la utilidad de los contenidos	4.5	4.7	4.6	4.6	4.8
10. Manifiesta entusiasmo por lo explicado	4.9	5.2	5.1	5.1	5.4
11. Ahora siento mayor interés por esta área	4.3	4.4	4.3	4.3	4.6
12. Responde con precisión	4.6	4.7	4.6	4.7	4.9
13. Mantuvo mi atención durante las clases	4.1	4.3	4.2	4.3	4.6
14. Explica de forma coherente	4.6	4.7	4.7	4.7	4.8
15. Atrae la atención del alumno sobre la materia	4.4	4.6	4.6	4.5	4.8
16. Explica con claridad	4.4	4.6	4.5	4.5	4.7
17. Estimula la participación	4.2	4.3	4.3	4.3	4.7
18. Se muestra competente en la materia	5.3	5.5	5.4	5.4	5.5

19. Adecuada preparación de la clase	5.3	5.5	5.4	5.4	5.5
20. Es constante en la asistencia a clase	5.9	6.1	6.0	6.0	6.0
21. Ajuste de lo explicado al programa	5.3	5.5	5.4	5.4	5.4
22. El examen permite reflejar lo aprendido	4.4	4.5	4.4	4.4	4.8
23. Examina sobre lo esencial	4.7	4.8	4.7	4.7	5.0
24. Las notas reflejan lo aprendido	4.2	4.3	4.2	4.3	4.7
25. La bibliografía sirve para preparar la materia	4.3	4.5	4.4	4.5	4.9
26. Me gustaría tenerlo de nuevo como profesor	4.1	4.1	4.0	4.1	4.4
27. Eficacia del profesor	4.6	4.8	4.7	4.7	5.0
28. Calificación comparativa con otros profesores	4.7	4.7	4.7	4.7	5.0
29. Calificación comparativa con otras asignaturas	4.5	4.6	4.6	4.6	4.7

Con respecto a la variable *turno* (Tabla 13), se constata la semejanza valorativa de los estudiantes de mañana y tarde, salvo ligerísimas variaciones, mientras que los alumnos de noche muestran valoraciones algo más altas en casi todos los ítems. Es pertinente señalar aquí que el número de estudiantes en el turno de noche es sustancialmente inferior al de los otros turnos y que, en su momento, concurrían características que lo diferenciaban de los otros dos grupos. Hoy ya no existe este turno.

Una de las críticas aparentemente más duras que se puede hacer al sistema de evaluación de la calidad docente mediante las opiniones de los alumnos es que éstos estarán influidos significativamente por las notas que reciben. El caso posiblemente más extremo lo encontraríamos en los *repetidores*. Si un alumno ha suspendido, cabe imaginar que evaluará más negativamente a su profesor que si hubiese aprobado. Este tipo de razonamiento era muy frecuente entre algunos de los profesores de las ciencias experimentales. Para dar cumplida respuesta a esta cuestión, separemos a todos los estudiantes que cumplieron el CEPUC intermedio en dos grupos: los repetidores y los no repetidores. Los resultados aparecen recogidos en la Tabla 14.

Éstos ponen de manifiesto que no existen diferencias relevantes entre sus valoraciones, dado el equilibrio global manifiesto cuando se tienen en cuenta todos y cada uno de los ítems: en unos los repetidores se muestran más generosos (1, 13, 16, 20, 21 y 27), en otros menos (4, 10, 11, 18, 22, 23, 24, 25 y 29) y en el resto hay plena coincidencia. Cabe destacar la evaluación algo más negativa en el ítem 22, referido al tipo de examen, y en el 24, referido a las notas, que resulta a todas luces comprensible, pero, sobre todo, lo más importante con relación a la cuestión planteada es que este aspecto no parece influir en lo más mínimo a la hora de realizar la evaluación de la calidad del profesor en cuanto tal. Tal vez estos datos constituyan uno de los apoyos empíricos más fuertes a la validez de este tipo de evaluación, pues nos indican que incluso aquellos aspectos que más pueden afectar emocionalmente a los alumnos están controlados cuando se les pide que valoren cualidades manifiestas del profesorado.

Otra de las cuestiones más frecuentes del profesorado es la referida al tiempo necesario para que los alumnos puedan emitir juicios con conocimiento de causa sobre la enseñanza recibida. Nos referimos a la duda en torno a si la *duración de la asignatura* puede incidir, para bien o para mal, en las evaluaciones de los alumnos. Los datos recogidos en la Tabla 14 ponen de manifiesto que ésta no es una variable que influya significativamente en las evaluaciones, globalmente consideradas, pese a las ligeras variaciones encontradas: las asignaturas anuales aparecen ligeramente mejor evaluadas, siendo las del segundo cuatrimestre las menos bien valoradas.

Tabla 14

La evaluación de la calidad docente de la Universidad Complutense en función de la variable tipo de alumno y duración de la asignatura

TIPO ALUMNO Y DURACIÓN ASIGNATURA	Repet.	Norepet.	Anual	1º cua.	2ºcua.
TOTAL EVALUACIONES	28.459	205.109	208.954	14.293	11.078
ÍTEMS (formato reducido)	Media	Media	Media	Media	Media
1. Su forma de enseñar facilita la comprensión	4.5	4.4	4.4	4.2	4.3
2. Muestra interés por las inquietudes del alumno	5.0	5.0	5.0	5.0	4.8
3. Motiva a preguntar sobre cualquier duda	4.5	4.5	4.6	4.4	4.4
4. Despierta interés por la asignatura	4.2	4.3	4.3	4.1	4.2
5. Se muestra accesible	5.0	5.0	5.0	5.1	4.9
6. Facilita la expresión de las ideas del alumno	4.6	4.6	4.7	4.6	4.5
7. Alude a las implicaciones de los enfoques	4.8	4.8	4.8	4.7	4.6
8. Establece interrelación entre los temas	4.8	4.8	4.8	4.7	4.6
9. Refleja la utilidad de los contenidos	4.6	4.6	4.6	4.6	4.6
10. Manifiesta entusiasmo por lo explicado	5.0	5.1	5.1	5.0	4.8
11. Ahora siento mayor interés por esta área	4.2	4.4	4.4	4.2	4.2
12. Responde con precisión	4.7	4.7	4.7	4.6	4.5
13. Mantuvo mi atención durante las clases	4.3	4.2	4.3	4.1	4.1
14. Explica de forma coherente	4.7	4.7	4.7	4.6	4.6
15. Atrae la atención del alumno sobre la materia	4.6	4.6	4.6	4.5	4.4
16. Explica con claridad	4.6	4.5	4.5	4.4	4.4
17. Estimula la participación	4.3	4.3	4.3	4.2	4.2
18. Se muestra competente en la materia	5.3	5.4	5.4	5.4	5.3
19. Adecuada preparación de la clase	5.4	5.4	5.4	5.4	5.3
20. Es constante en la asistencia a clase	6.1	6.0	6.0	6.1	6.0
21. Ajuste de lo explicado al programa	5.5	5.4	5.4	5.5	5.4
22. El examen permite reflejar lo aprendido	4.2	4.5	4.4	4.4	4.3
23. Examina sobre lo esencial	4.6	4.7	4.7	4.7	4.6
24. Las notas reflejan lo aprendido	4.1	4.3	4.3	4.1	4.0
25. La bibliografía sirve para preparar la materia	4.4	4.5	4.5	4.5	4.4
26. Me gustaría tenerlo de nuevo como profesor	4.1	4.1	4.1	4.0	4.0
27. Eficacia del profesor	4.8	4.7	4.8	4.6	4.6
28. Calificación comparativa con otros profesores	4.7	4.7	4.7	4.6	4.5
29. Calificación comparativa con otras asignaturas	4.5	4.6	4.6	4.4	4.4

A la luz de todos estos resultados referidos a algunas de las variables que los profesores pensaban en un comienzo que podrían tener cierta incidencia en la evaluación de la calidad docente, se puede concluir que esa supuesta incidencia en todos los casos es mínima, de forma que la valoración global de los alumnos es plenamente válida. Además, se constata que el perfil de los puntos fuertes y débiles en todos los casos es prácticamente idéntico, con independencia de los desgloses que se lleven a cabo, en función de las distintas variables de interés: sexo, turno, curso, etcétera.

CAPÍTULO QUINTO
SÍNTESIS DE LO CONCEPTUAL Y LO
OPERATIVO

1. La cultura de la evaluación y la utilización de los resultados de la evaluación

Cualquier cultura de la evaluación ha de asentarse sobre dos pilares básicos. Por un lado, el principio universal constatable de que los humanos, por el hecho de serlo, estamos condenados a ser permanentes objetos y sujetos de evaluación. No hay escapatoria posible. Esto vale tanto para los docentes como para los discentes, para las autoridades académicas así como para los representantes políticos, encargados del diseño y desarrollo de las normativas legales referidas a la educación e, incluso, para el personal de administración y servicios que ejerce sus funciones dentro de los ámbitos educativos. Por otro lado, es igualmente constatable que toda educación es un acto propositivo: queremos lograr algo con dicho acto (conocimientos, habilidades, competencias, sistemas de creencias). En estricta lógica eso conlleva el que necesariamente hayamos de contar con la evaluación, bien de tipo informal (lo que ha ocurrido a lo largo de los siglos) o bien, a partir del siglo XX, con sistemas valorativos que gozan de apoyo científico y técnico. Tampoco aquí hay escapatoria posible. Querámoslo o no, las personas valoramos, con mayor o menor rigor, los currículos, los contextos, las interacciones entre profesores, alumnos y padres, los logros obtenidos por unos y otros o las actuaciones de los gestores educativos. *De ambos principios esenciales surge una clara conclusión: más vale emplear una evaluación avalada científicamente, pues es en el mundo académico donde estamos, que recurrir a las meras impresiones de los distintos agentes que participan en los diferentes contextos de enseñanza/aprendizaje o, lo que sería todavía peor, a las impresiones de ciertas autoridades educativas, en ocasiones demasiado alejadas del mundo en el que se produce el proceso docente/discente real.*

Tirios y troyanos no tienen hoy en día más remedio que aceptar estos postulados. Son, de hecho, los que se están aplicando en multitud de parcelas del mundo que nos rodea. Si no se llevaran a cabo las pertinentes evaluaciones difícilmente se podrían perfeccionar los productos que utilizamos y consumimos diariamente (desde los coches hasta la vivienda, pasando por las máquinas/instrumentos de diagnóstico y tratamiento médico). Por eso, no nos ha de extrañar que esta poderosa herramienta –la evaluación– goce en nuestros días de la correspondiente “institucionalización académica”, es decir, haya entrado en la vía de la ciencia. Esto no significa que, a comienzos del siglo XXI, sepamos a ciencia cierta cuál es la función del evaluador, pues como vimos en el primer capítulo, puede haber, y de hecho hay, considerables discrepancias, aunque en todos los casos se abogue por que los “conocimientos científicos” avalen el informe evaluador.

No deja de ser curioso que pese a que muy pocos, si es que alguno, se atreven a poner en duda estos postulados sobre los que se asienta la evaluación, en el terreno de la enseñanza formal los informes de evaluación suelen dormir el sueño de los justos en numerosos casos. En un principio, resulta difícil entender esta situación. Se quería tener conocimiento de los puntos fuertes y débiles de lo evaluado (un programa, un curso, un currículo, profesores o alumnos), se ha realizado una inversión de recursos humanos y materiales, se cuenta con uno o varios informes técnicos y, al final, se opta por dejar las cosas como están. ¿Por qué esta situación resulta tan frecuente, al menos en el campo educativo, aunque en modo alguno sea éste el único ámbito en el que tiene lugar (Patton, 1997; Weiss, 1988)?

La explicación ha de buscarse en el omnímodo poder de los intereses de quienes están llamados a tomar las decisiones últimas. Un partido que ha propuesto una reforma

con las mejores intenciones del mundo (objetivos cargados de connotaciones positivas: la calidad o excelencia de los procesos/productos educativos), cuando encuentra que los datos de la evaluación no apoyan precisamente lo prometido en la campaña electoral (tal vez ha empeorado la situación o, simplemente, no se han producido cambios significativos en la dirección prevista), hace todo lo posible para que estos resultados no vean la luz, pues sin duda servirán de munición para los disparos de la oposición. Si sustituimos partido político por rector, decano, director de centro, profesor o cualquier otro agente de la enseñanza, comprenderemos que todos ellos estén fuertemente tentados a seguir el mismo tipo de comportamiento: *la ocultación o el olvido de los resultados cuando no les son del todo favorables*. En este contexto, la evaluación ha de pagar caro por una culpa que en modo alguno le corresponde. Es víctima de los intereses cruzados de quienes en ese momento detentan el poder. No es fácil salir de este extendido atolladero. De ahí la ventana abierta que supone la denominada cultura de la evaluación. Ésta requiere un diálogo, una comprensión, una adopción de perspectivas por parte de todos los miembros que participan en el complejo proceso de enseñanza y aprendizaje, de forma que los resultados de la evaluación no supongan un arma arrojada contra los posibles adversarios ideológicos o políticos. Ardua tarea ésta para desarrollar a lo largo de todo el siglo XXI (Leviton, 2003). De momento, cabría recomendar que no se inicie ningún proceso evaluador (se podría convertir la evaluación en una auténtica plaga que ocasionase un odio visceral a la misma) a menos que estén aseguradas las pertinentes decisiones ulteriores, siempre fundadas en los datos recogidos de forma válida y fiable. Hemos de ser conscientes de que no sólo no tiene ningún sentido, sino que es claramente contraproducente (se consumen innecesariamente recursos humanos y materiales, se crea desconfianza y se desmotiva a los participantes) realizar evaluaciones, bien asentadas científicamente, cuyos resultados no vayan a servir para nada. Muy probablemente sea ésta una de las mayores encrucijadas con las que se ha topado la evaluación hasta estos momentos, pues siendo así que ha logrado ya el estatus científico tan perseguido a lo largo del siglo pasado – paso de la evaluación informal a la formal-, ahora ha de hacer frente a la siguiente omnipresente pregunta: *después de los resultados de la evaluación, ¿qué?*

Con el fin de ir limando los razonamientos a los que a veces se recurre para no hacer caso de los resultados de la evaluación, los evaluadores han de poner sumo empeño en tener en cuenta todos y cada uno de los matices esenciales que garanticen que el proceso evaluador ha seguido los pasos de rigor (buena fundamentación científica y falta de sesgos apreciables) en los resultados obtenidos, a la par que encuadrar todo ello dentro de una imprescindible cultura de la evaluación.

2. El proceso evaluador

Mal se comienza, y muy probablemente peor se acabe, si se asume que la evaluación es tarea sencilla y que basta con poner buena voluntad para que se llegue al puerto anhelado. Ello significaría un claro menosprecio al arduo trabajo desarrollado a lo largo de todo un siglo para hacer que podamos pasar de la evaluación informal a la formal. Cualquier proceso evaluador requiere una atención especial en tres momentos cruciales:

- el inicial;
- el del desarrollo e implantación de las evaluaciones;

- el de la reflexión académica sobre el informe evaluador.

En el inicio, se impone una clarificación y un compromiso. Es imprescindible concretar (aclarar al máximo) cuáles son (o eran) los objetivos de quien es el responsable de un determinado acto propositivo de carácter educativo (currículo, curso, programa, titulación), a fin de que desaparezcan, en lo posible, las ambigüedades y las palabras sin apenas denotaciones pero con excesivas connotaciones (excelencia, calidad, creatividad, desarrollo óptimo, autorrealización, valores, etcétera). Al encontrarnos con estas palabras “comodín”, se debiera hacer un considerable esfuerzo por pedir ejemplos (indicadores concretos) a los políticos y responsables académicos, dado que en buena parte de los casos sólo se nos ofrece un puro nominalismo. Sin esta clarificación, más vale no seguir adelante, pues la supuesta *evaluación científica corre peligro de que se convierta en un mero instrumento técnico de justificación ideológica*.

Tras la clarificación, el compromiso. Si los responsables de poner en marcha el proceso evaluador no están dispuestos a tener en cuenta los resultados del informe y actuar en consecuencia, tampoco merecería la pena que se continuase con la evaluación. Sería una auténtica pérdida de tiempo y se minaría la confianza de todas aquellas personas que han dedicado su tiempo y esfuerzo a la recogida de los datos. De nuevo, pues, *la evaluación sólo tiene sentido dentro del marco de la cultura de la evaluación, en tanto ésta implica una voluntad firme de aceptar el veredicto (positivo o negativo, favorable o desfavorable) del informe evaluador*.

Suponiendo, lo que ya es mucho suponer, que se ha dejado de lado el puro nominalismo y que se ha conseguido el compromiso de actuar según los resultados de la evaluación, el paso siguiente es lograr un buen diseño de evaluación formal. Generalmente, esto sólo se obtiene –condición necesaria aunque no suficiente- cuando se han asimilado los *avances que han tenido lugar a lo largo de todo el siglo XX en lo tocante a los distintos procedimientos y enfoques de la evaluación*: se tienen en cuenta sus ventajas y desventajas, antes de decidirse por un determinado enfoque, o por una síntesis o integración de varios de ellos. En principio, sea cual sea la orientación adoptada, parece prescriptivo que se realice una evaluación inicial, gracias a la cual es posible conocer “científicamente” la línea de base o partida y establecer una jerarquía de los asuntos a evaluar, a fin de no consumir recursos humanos y económicos en asuntos de escasa relevancia. Estos primeros datos nos han de servir de condicionantes básicos para una evaluación ajustada al contexto, tanto de procesos como de productos. Hemos de ser conscientes de que el contexto (monto económico disponible, personal con el que contar, extensión y profundidad deseada en la recogida de los datos, tipo de decisiones que se piensan tomar a raíz de los resultados) condiciona en buena parte el tipo de evaluación a emplear, sabiendo que no existe la evaluación perfecta, válida y fiable en todo momento y circunstancia.

Parece oportuno tener en cuenta aquí que no es bueno dejarse llevar por *falsas dicotomías* que, a poco que se analicen, no tienen excesivo sentido. Tal vez la que destaca por encima de las demás sea la de lo cualitativo frente a lo cuantitativo. Uno puede inclinarse por una metodología más cualitativa o más cuantitativa pero, *a priori*, no se puede decir que una sea mejor que la otra por el hecho de ser cuantitativa o cualitativa, puesto que la calidad de la evaluación reside más bien en la capacitación del evaluador. De hecho, es fácil encontrar malas evaluaciones cuantitativas o cualitativas,

al igual que ocurre con las buenas. Luego su calidad reside, entre otros factores, en el diseño e interpretación de los datos, funciones que le corresponden única y exclusivamente al evaluador. Algo semejante cabría indicar con respecto a la contraposición tan frecuente entre modelos (análisis de sistemas, libre de metas, cuasi-legal), si bien es cierto que, en este caso, algunos de ellos en sí mismos considerados, pueden gozar de un número mayor de ventajas que otros, pero de nuevo es la capacitación del evaluador la variable que tal vez explique mayor proporción de varianza de la calidad de la evaluación.

Al hablar de la calidad de la evaluación –de productos y de procesos, más que de procesos frente a productos- hemos de referirnos necesariamente a la *metaevaluación*. Como vimos, ésta nos proporciona una serie de criterios consensuados por la comunidad de evaluadores, e institucionalizados mediante las distintas asociaciones nacionales e internacionales de evaluación, gracias a los cuales es posible garantizar la fiabilidad y validez del informe científico-técnico de evaluación. Estos criterios, afortunadamente, tratan de superar tanto las falsas dicotomías como el enfrentamiento entre los distintos modelos. La dura realidad nos indica que con demasiada frecuencia lo realizado, por un lado, y lo que se considera necesario/ideal desde la metaevaluación, por otro, no coinciden plenamente. Como resulta obvio, no siempre es culpa de la capacitación del evaluador. Los intereses espurios de los distintos agentes (sobre todo de los que ocupan puestos de poder) que participan en el contexto de aprendizaje y enseñanza suelen condicionar considerablemente la calidad de todo el proceso evaluador.

En este contexto, cada vez con mayor fuerza a medida que nos vamos adentrando en el siglo XXI, se está reclamando un esfuerzo para llevar a cabo las *evaluaciones de seguimiento*, pese a las enormes dificultades que éstas entrañan tanto desde el punto de los recursos humanos como económicos. Sin embargo, sólo ellas nos posibilitan disminuir la probabilidad de errores sobre el valor de los diferentes actos propositivos, errores que como ya indiqué (capítulo segundo) acechan a cualquier tipo de evaluación cuando ésta tiene lugar nada más terminar la actividad evaluada.

3. Los diversos procedimientos de evaluación

Hay pocas dudas de que el avance científico en asuntos de evaluación a lo largo del pasado siglo ha sido enorme. Hasta tal punto es así que hoy en día contamos con una *tecnología lo suficientemente poderosa* (se puede medir/valorar casi todo con un aceptable rigor) como para poder ayudar a perfeccionar considerablemente la calidad de la docencia, en cualquier parte del mundo. Ahora bien, la evaluación no se puede identificar sin más con medida, por muy válida y fiable que ésta sea. La evaluación se adentra plenamente en el ámbito de los valores y, en no pocas ocasiones, ambos mundos –el de la medida y el de los valores- no caminan juntos. Si a esto se añade que la evaluación necesariamente supone una toma de postura frente a intereses políticos de muy distinto signo o cariz, tenemos los ingredientes básicos que explican la enorme distancia entre lo que realmente podría aportar la evaluación en aras de la mejora docente y lo que de hecho aporta, víctima ésta de múltiples intereses espurios por parte de quienes detentan el poder social y académico, aparte, como es obvio, de las posibles deficiencias en las competencias del propio evaluador.

Dentro del marco estrictamente académico, es necesario tener en cuenta que no existe en la actualidad un único enfoque fiable y válido frente a los demás. Cada una de las perspectivas que hemos visto o insinuado en el capítulo segundo nos ofrece sus puntos positivos y negativos, lo que en modo alguno quiere decir que todas gocen de semejante respaldo científico. De hecho, un análisis de las *cuestiones típicas* de los distintos *modelos*, aunque sea somero, nos permite llegar a la conclusión de que la mayoría, si no todas ellas, debieran ser tenidas en cuenta por cualquier buen evaluador en el ejercicio de su profesión. Qué duda cabe de que, de una u otra manera, los objetivos han de servirnos de orientación inicial y final en cualquier proceso evaluador (*modelo de objetivos conductuales*). Ahora bien, en nuestro siglo, parece oportuno que el logro de esos objetivos venga condicionado por el matiz de la *eficiencia*, es decir, el evaluador debiera preocuparse por saber a ciencia cierta si se puede lograr lo mismo más económicamente, con una mejor relación de costes beneficios o efectividad (*modelo de análisis de sistemas*). Hasta aquí, el evaluador se habría centrado fundamentalmente en el producto final, siendo así que parece más completo detenerse también en las distintas partes de ese producto que hasta ahora se había considerado como un todo (*modelo de toma de decisiones*). Si se tienen en cuenta estos matices propuestos desde este modelo, el evaluador estaría bien preparado para afrontar los desafíos de la *calidad total*. Pero el proceso no acaba aquí. El viaje ha de continuar.

De momento, hemos estado muy centrados en los objetivos previstos a priori, de forma que éstos han condicionado casi toda la labor del evaluador. Parece llegado el momento de subir el siguiente peldaño: el que hace referencia a todos los efectos producidos tras el acto propositivo (*modelo libre de metas*). Esto son ya palabras mayores. Por eso, no todos los evaluadores se atreven a dar este paso. Les parece suficiente constatar el grado en que se han cumplido los objetivos de forma rigurosa. Incluso no les importa mostrar el precio (en recursos materiales y humanos) pagado por determinado nivel de logro. También están dispuestos a desglosar el producto final en sus componentes o partes esenciales. Pero investigar todos, lo que se dice todos los efectos, eso es harina de otro costal. Sin embargo, es la subida a este nivel la que nos permite vislumbrar el futuro. Cada vez más se irá exigiendo a los evaluadores que se tomen en serio el análisis de todos los efectos: los positivos y los negativos, los previstos y los imprevistos, los deseados y los no queridos. En este contexto, se pueden enmarcar buena parte de las *aportaciones del resto de modelos* (*crítica de arte, revisión profesional, cuasi-legal, o estudios de casos*): se necesita contar, pues, con profesionales (con los amigos y hasta con los enemigos, con los que comparten mi enfoque y con los que manifiestamente están en contra) e incluso con los paraprofesionales, pues ellos nos pueden proporcionar visiones o perspectivas complementarias a la estrictamente profesional.

Lo dicho anteriormente tal vez explique el porqué en un determinado momento histórico (las décadas de los 70 y los 80 básicamente) parecían resultar de utilidad las *taxonomías de enfoques*. Hacían justicia al desarrollo y productividad variopinta de un campo —el de la evaluación formal— en permanente ebullición. Hoy en día los ánimos se hallan más sosegados —es posible una reflexión en profundidad de las semejanzas y diferencias de los distintos enfoques—, por lo que los evaluadores, ya maduros, tratan de conciliar lo que en un principio parecía contrapuesto. Son conscientes de que la unión de los puntos fuertes de cada enfoque puede resultar ideal para la práctica de la evaluación y para la credibilidad de los evaluadores. La renuncia a una supuesta pureza teórica —el seguimiento estricto de un modelo— se ve recompensada con creces por el

éxito con que le refuerza la práctica, pues ésta exige mentes abiertas en donde quepa parte de la complejidad de la realidad. Así, desde la amplitud de elección de enfoques de que goza en un principio cualquier evaluador, una fuerza centrípeta parece guiarlo a un lugar en donde se encuentra el denominador común de principios básicos derivados de estos modelos, *conditio sine qua non* de cualquier buen profesional de la evaluación.

4. Los enfoques integradores

El modelo CIPP, al que se hizo alusión en el segundo capítulo, es un claro ejemplo verificador de lo que se acaba de exponer. De ahí, muy probablemente, la buena acogida que desde un principio tuvo este modelo integrador por una gran parte de evaluadores y que le ha supuesto una garantía de continuidad todavía en nuestros días. Ahora bien, no cualquier tipo de integración hace bueno al modelo, pese a que algunos así lo crean, ni la misma supone que todos los modelos cobijados bajo el paraguas integrador gocen de idéntica protección frente a posibles sesgos. Me voy a servir de dos enfoques integradores ya vistos –el de la *carpeta docente* y el de la *evaluación circular*– para poner de manifiesto las importantes diferencias subyacentes entre ellos, pese a que en ambos se advierte esa voluntad de incorporar los aspectos más positivos de otros enfoques (semejanzas).

Dentro de la variedad de portafolios es difícil encontrar alguno que no incluya la necesidad de tener en cuenta, y por tanto archivar, las opiniones de los alumnos como uno de los elementos más esenciales. Por su parte, tal cual se vio, el modelo de la evaluación circular implica la recogida continuada de las opiniones de los alumnos, que obliga tanto al profesor como a los distintos responsables académicos a archivarlas en una carpeta con el fin de ir viendo empíricamente hasta qué punto mejora o empeora su docencia. En los dos enfoques se requiere la reflexión e investigación en la acción, a partir de los datos obtenidos. También en los dos se parte de que el profesor debe ser activo ante la información recibida e, igualmente, desde los dos se supone que los datos compilados han de servir ante todo y sobre todo para el desarrollo profesional y para el continuo perfeccionamiento docente, sin que esto implique excluir el que los resultados puedan encontrar también algún tipo de función sumativa. La lista de semejanzas se podría alargar considerablemente al pensar que los dos asumen buena parte de los planteamientos constructivistas, tanto por lo que se refiere al profesor como en lo tocante a los alumnos, que incorporan la autoevaluación, que tienen en cuenta las opiniones de los responsables académicos y, por supuesto, que lo que realiza el profesor en el aula es primordial a la hora de forjarse una imagen válida y fiable de la enseñanza del profesor. Si se sigue esta línea de argumentación, ¿uno no estaría tentado de concluir que *ambos modelos son intercambiables pues los dos parecen pertenecer a la familia de los enfoques integradores?*

Creo que este es un error que no debiera cometerse. Pese a las aparentes semejanzas vistas, pese a ese anhelado denominador común formado por las ventajas que se han ido descubriendo al poner en funcionamiento los “enfoques más puros” –los que surgen con el objetivo de contraponerse a los existentes: análisis de sistemas, estudios de casos, entre otros– al que sin duda alguna estos dos modelos actuales aspiran –como también aspiran otros muchos: el de la calidad total, el de los 360 grados (Wilkerson, Manatt, Rogers y Maughan, 2000), por ejemplo–, en modo alguno cabe afirmar que sean intercambiables. ¿Por qué? *Porque, pese a las semejanzas en las intenciones (en aquellos aspectos más teóricos), las grandes y fundamentales*

diferencias se dan en el terreno de la práctica. Así, mientras que en un caso se habla de las opiniones de los alumnos sin más, o a lo sumo se dice que sean válidas y fiables, en el otro se proporcionan unos instrumentos de evaluación, se los compara con otros ya existentes y se llevan a cabo metaevaluaciones que nos desvelan multitud de matices de sumo interés, como tuvimos la oportunidad de comprobar en el capítulo cuarto. Incluso si nos adentramos en un aspecto esencial para ambos modelos –el cruce de informaciones-, comprobamos que en un caso se indica única y exclusivamente que éste es bueno (esa es la finalidad última del portafolios), pero quedamos un poco a oscuras cuando nos interrogamos sobre el cómo –a lo sumo se nos dice que se deja a la creatividad del profesorado-, mientras que en el otro caso, a la par que se afirma que el cruce de informaciones es el aspecto clave del modelo se nos da también un modo concreto y específico de hacerlo tras, una vez más, haberlo comparado en la práctica con otras formas o modos concretos de proporcionar ese cruce de información. Siguiendo con las desemejanzas en aspectos sustanciales, constatamos que mientras en un caso casi toda la responsabilidad recae en el profesorado, en el otro es plena y democráticamente compartida: cada uno de los agentes que participa en cada contexto de aprendizaje y enseñanza es evaluador y evaluado al mismo tiempo, según las competencias y las responsabilidades de cada uno. El profesor ha de cumplir su cometido y será evaluado por ello, pero otros han de realizar lo más correctamente posible el suyo, por lo que igualmente serán evaluados. La mejora de la calidad implica que todos, eso sí, en diverso grado, han de participar en ese delicado y complejo proceso y, para ello, la evaluación de todos y cada uno es esencial –su cruce-, a fin de que puedan reflexionar y compartir experiencias encaminadas a esa calidad total o excelencia docente, definida de forma operativa y no meramente de forma teórica.

A la luz de lo dicho y justificado, podemos concluir que si bien los modelos integradores gozan hoy en día de un merecido prestigio, pues suponen un intento de integración de las fortalezas de los modelos más puros, sin embargo, en modo alguno son todos ellos intercambiables, por lo que es necesario un análisis muy fino y bastante exhaustivo antes de inclinarnos por uno u otro a la hora de ponerlos en funcionamiento. A esto ha de añadirse, que la posible bondad teórica de cada uno de estos modelos ha de compararse teniendo en cuenta las demandas del contexto (recursos materiales y humanos; tipo de experiencias previas; finalidad de la evaluación –formativa o sumativa-; existencia o ausencia de una determinada cultura de la evaluación, etcétera-), ya que éstas pueden ser con toda justicia condicionantes a la hora de la elección de un determinado enfoque.

5. La evaluación, compañera de viaje de cualquier profesor

En el capítulo tercero vimos que, querámoslo o no, un profesor pasa por diversas etapas en su desarrollo profesional. Cada una de ellas requeriría un proceso específico de evaluación. Luego, *la evaluación se va a convertir en una compañera necesaria de ese viaje profesional.*

Las múltiples evaluaciones realizadas a lo largo de una determinada carrera –las notas- no parecen ser suficientes, por buen expediente académico que uno haya podido lograr, para garantizar el éxito como docente en cualquiera de los niveles a los que nos podamos referir: primaria, secundaria o universitario. Si bien el *título* es sin duda una condición necesaria para la impartición de unos determinados contenidos, en modo alguno es suficiente, tal como pudimos comprobar empíricamente en las evaluaciones

del CAP, según comentarios de los propios licenciados en diferentes Titulaciones. Ante esta situación, los distintos países han elaborado muy diversos programas con el objetivo de enseñar básicamente el cómo más que el qué, pues ésta parecía ser la deficiencia fundamental en casi todas las carreras universitarias. Las evaluaciones de la mayoría de estos programas, al menos en los Estados Unidos, no fueron precisamente positivas, dado que se reclamaba su supresión. De cara al inmediato futuro (la implantación de un máster como sustituto del CAP) se requiere, por tanto, tener en cuenta estos datos –paradojas-, pues mientras se reclaman como necesarios esos programas por parte de los interesados, tras cursarlos, ellos mismos los consideran, al menos en parte, una pérdida de tiempo.

En esta situación, no precisamente ideal, han de hacer frente a la *selección*. Aquí hablamos de otro tipo de evaluación bien distinto. Si seguimos la perspectiva scriveniana, el mérito, el valor y la utilidad tendrían que ser los tres pilares sobre los que debiera descansar esta clase de valoración. Incluso dentro de este enfoque, ya gozaríamos de múltiples y bien diferenciadas técnicas para la recogida de la información que con vistas a este fin –el de la selección-, necesitamos. El avance logrado a lo largo del siglo pasado en la validación de estas técnicas es considerable, si bien es posible comprobar que, a veces, los avances científicos y técnicos van por un lado y la práctica de la evaluación por otro bien distinto, debido a intereses de muy diverso tipo.

Seleccionado el profesor, éste ha de superar las valoraciones de los *primeros años de ejercicio profesional*. También aquí podemos decir que se cuenta con procedimientos suficientes como para que la evaluación pueda ser considerada formal y no meramente informal, aunque la realidad puede mostrarnos considerables distancias entre lo posible y lo que acontece de hecho. A partir de aquí y hasta la jubilación, *continuas evaluaciones*, bien encaminadas a la mejora docente y al desarrollo profesional, bien a subidas salariales o promociones o incluso al cese de la contratación, van a tener lugar, pues, como se ha ido poniendo de manifiesto a lo largo del libro, la evaluación es consustancial a los humanos por el hecho de estar guiados por actos propositivos. Mucho más cuando estos actos tienen lugar dentro del marco del saber científico que por sí mismo requerirá que la evaluación sea también científica.

6. La evaluación del profesor en el contexto de las demás evaluaciones

Como vimos en el capítulo primero al realizar un análisis tanto de la cultura de la evaluación como del ámbito educativo, estos constructos en modo alguno pueden entenderse si nos limitamos al examen y a la valoración de uno de sus agentes –el profesorado-, por importante que éste pueda ser, como de hecho lo es (Ellett y Teddlie, 2003). De ahí que la evaluación sobre este colectivo haya de acompañar necesariamente a las otras evaluaciones realizadas –volvemos a la Figura 1- sobre cada uno de los componentes que constituyen los pilares fundamentales del mundo de la educación (Centra y Potter, 1980).

Las *evaluaciones de los alumnos*, desde la perspectiva aquí adoptada, no han de limitarse a las notas, sino tener en cuenta al menos tantas facetas como hemos puesto de manifiesto en el caso del profesor: por supuesto, conocimientos (rendimiento), pero también capacidades (habilidades) y su desglose en competencias, sin infravalorar los valores, las creencias y las actitudes. No debiera olvidarse que la calidad o excelencia en

el ámbito educativo sólo se produce cuando contamos con buenos profesores y buenos alumnos (conoceríamos su bondad precisamente gracias a la evaluación). Si esto es verdad, las omnipresentes falsas dicotomías (los profesores creen que la falta de calidad se debe a los alumnos: a su falta de capacidad, de motivación, o de interés, mientras que los alumnos piensan lo contrario: de hecho, ellos siempre aprueban y son los profesores los que les suspenden, casi siempre injustamente) debieran dar paso necesariamente a cuádruples tipologías:

- en un cuadrante nos encontraríamos con buenos profesores y alumnos (buena calidad del proceso/producto de enseñanza/aprendizaje);
- en el cuadrante opuesto con malos profesores y malos alumnos (mala calidad educativa);
- en los dos restantes: con buenos profesores y malos alumnos y con buenos alumnos y malos profesores (relativa calidad educativa en ambos casos, aunque con matizaciones diferenciadoras entre ellos).

Aquí ya estamos vislumbrando la *circularidad*, que nos indica que lo que hace el alumnado condiciona al profesorado y lo que este colectivo ejecute va a tener repercusiones en aquél, aunque con valores ponderativos distintos en cada caso. Por eso las flechas son bidireccionales en la Figura 1. La calidad del proceso de enseñanza/aprendizaje es cosa al menos de dos. Luego no puede haber un miembro activo y otro pasivo, como en el modelo clásico. Los dos agentes han de ser activos y los dos han de compartir unos conocimientos y valores que ya existen, a la par que intentar conseguir nuevas aportaciones o crear nuevos conocimientos. La evaluación de ambos grupos, por tanto, se complementa.

Ahora bien, como se señalaba en la Figura 1, el binomio ha de dar paso al trío (triángulo) de componentes básicos: el tercero es el *currículo*, entendido en su sentido más amplio (explícito, oculto, formal, no formal e informal). Éste también, pues, ha de ser evaluado (McCormick y James, 1983; Lewy, 1991). Su mayor o menor grado de bondad se hará sentir en esa *calidad total* a la que estamos aspirando. De nuevo aparece la bidireccionalidad para indicarnos que el currículo va a incidir tanto en alumnos como en profesores y que, a su vez, ambos colectivos van a condicionar el tipo de desarrollo de aquél. Uno de los lazos más importantes de unión de estos componentes y de su interacción lo constituye precisamente la evaluación.

Pero al detallar la especificidad del ámbito educativo se nos puso de manifiesto que el comportamiento de otros agentes (los padres, el personal de administración y servicios, las autoridades académicas o los responsables políticos) tenía que ser considerado, en la medida en que nuestro propósito fuese el de la comprensión científica del grado de excelencia global de todo un proceso/producto de enseñanza/aprendizaje. La unión y cruce de datos evaluativos de todos estos miembros se constituye, por consiguiente, en pieza clave para, primero, entender lo que sucede en un momento determinado y, segundo, proceder a mejorarlo. Luego, la evaluación de todos ellos se hace imprescindible, desde cualquier punto de vista.

Una última unidad de análisis, una de las más abarcadoras si no la que más, sería el centro (Hopkins, 1989). En él se encierran todas las evaluaciones necesariamente

parciales a las que hemos hecho alusión hasta estos momentos. El nivel de complejidad alcanzado, como resulta obvio, es máximo.

Pues bien, ¿el modelo de evaluación circular nos podría seguir siendo útil más allá de la evaluación del profesorado? Yo creo que sí. Entre otras razones, por las siguientes:

- Está diseñado para explicar el comportamiento de *sistemas complejos*, como de hecho lo es el educativo (véase Figura 1). No sólo tiene en cuenta las unidades de análisis clásicas (profesores y alumnos, básicamente), sino al resto de participantes (padres, autoridades académicas y políticas, personal de administración y servicios) y a los distintos contextos en donde han de ejercer sus funciones cada una de estas personas (microsistemas, mesosistemas y macrosistemas).
- La *bidireccionalidad* es ominipresente frente a la unidireccionalidad de los enfoques más clásicos. Aquella tiene lugar tanto entre individuos (profesor/alumno, alumnos entre sí –aprendizaje cooperativo-) como entre los diversos agentes y los diferentes contextos en los que se desenvuelven (educación formal, no formal e informal).
- La *circularidad* es pieza clave a la hora de explicar científicamente los distintos grados de excelencia en todos y cada uno de los niveles o unidades de análisis: alumnado, profesorado, currículo, centro,... De ahí la necesidad del cruce de los distintos resultados evaluativos procedentes de todos y cada uno de los componentes de cualquier sistema educativo.
- Los resultados obtenidos hasta el momento ponen de manifiesto que la contraposición tan de moda entre los planteamientos cualitativos y cuantitativos, y su dulcificación mediante un lenguaje de lo políticamente correcto –complementariedad-, no tienen mucho sentido. Durante más de dos décadas de aplicación del modelo de evaluación circular no he sido capaz de encontrar la más mínima contradicción en las conclusiones derivadas de uno u otro planteamiento. Sí he constatado, en cambio, *resultados muy dispares obtenidos por malos frente a buenos evaluadores de uno u otro signo*.
- Es, o pretende ser, un *enfoque integrador*. La evolución de la evaluación formal a lo largo de todo el siglo XX nos ha permitido ir asimilando continuamente aspectos relevantes de la misma. Éstos se han ido incorporando sin grandes problemas a la estructura inicial del enfoque (recuérdese, por ejemplo, la incorporación reciente de la autoevaluación).
- Estechamente unida a la característica anterior está la de su *flexibilidad*. Ya tuvimos oportunidad de ver cómo los distintos instrumentos de evaluación (CEPUC, CECA,...) han ido sufriendo modificaciones (distintas versiones) con el fin de lograr una acomodación máxima a las exigencias contextuales.

- *Asume los distintos tipos de patrones o normas del evaluador profesional*, que han sido consensuados por aquellos comités que gozan en nuestros días de un mayor prestigio internacional (por ejemplo, el Joint Committee on Standards for Educational Evaluation), para la evaluación de programas y miembros que participan en los distintos ámbitos de enseñanza y aprendizaje. Incluso se tienen en cuenta aquellos principios o patrones rectores que habiendo surgido del análisis de contextos virtuales (evaluación y enseñanza en línea), son de aplicación también en los contextos de enseñanza y evaluación presenciales (McCormack, 2005)

Obviamente, como cabía esperar, el autor de un enfoque (el del modelo de evaluación circular) ha de tratar de poner de manifiesto aquellos aspectos que cree proporcionan un mayor valor científico a su planteamiento. Pero, como ya indiqué al principio del libro que aquí acaba, el lector –ahora evaluador- tiene la última palabra, que precisamente puede ser ya palabra de “auténtico” evaluador o, si realmente se ha aplicado a fondo, de metaevaluador (prometo que todavía no he elaborado instrumentos suficientemente válidos y fiables como para distinguir a los unos de los otros, aunque con el tiempo, y en función de la retroalimentación recibida, tal vez no esté muy lejano ese día).

Epílogo metaevaluativo: el libro que no quise escribir

Una vez que el lector ha podido juzgar lo escrito -el libro-, me parece conveniente que sepa lo que no quise escribir, por ser consecuente con mi propio planteamiento: el del profesional que ha de seleccionar reflexivamente aquello que está científicamente mejor fundamentado. En una sociedad, como la nuestra, del conocimiento y de la información, de fácil acceso a Internet y a las más diversas bases de datos (MEDLINE, PsycINFO, ERIC, etcétera), a las revistas y libros electrónicos (e-journals y e-books, respectivamente), a las bibliotecas y librerías virtuales (amazon.com, por ejemplo), a buscadores de tipo más general (sea, entre otros, google.com) o más específico (scholar.google.com), no me pareció lo más correcto realizar una búsqueda exhaustiva de todo lo concerniente a la evaluación del profesorado para, después, ofrecerle al lector ese cúmulo inmenso de datos (¿conocimientos?). Sólo yuxtaponiendo lo encontrado habría que haber escrito no uno sino cientos de libros. De hecho, como veremos a continuación, los trabajos publicados sobre los distintos capítulos del libro, nos obligarían a hablar, sin exageración, de miles.

Desde mi punto de vista, hay hoy un exceso de información sobre casi todo y muy especialmente sobre la evaluación y más concretamente sobre la evaluación del profesorado. De hecho, si uno intentase asimilar todo lo que sería capaz de encontrar utilizando los distintos medios arriba indicados, tal vez feneciese en el intento. De ahí que, como psicólogo, no haya querido, preventivamente, que tenga que pasar nadie por una unidad clínica, tras el posible empacho lector. Ahora bien, ¿qué ocurriría si con un cierto hilo de Ariadna, como el que se ha tratado de ofrecer en el libro, alguien se interesase en complementar o matizar lo allí expuesto mediante estos nuevos medios informatizados? Esta alternativa, *a posteriori* y no *a priori*, es la que me parece más fructífera a la par que agradable de cara al futuro (pero con matices, como comprobaremos de inmediato).

Por eso, bien como despedida o bien como inicio de futuros encuentros, invito al lector a que, fijándose en algo que le haya llamado más la atención, proceda a llevar a cabo alguna búsqueda como la sugerida. Ejecutemos un ejemplo. Imaginemos que nuestro interés se centra en la carpeta docente. Me voy a google.com e introduzco la palabra inglesa *portfolio*. En unos segundos ha encontrado más de 250 millones de lugares en donde se habla de portafolios, en sus múltiples y variados sentidos. ¿Seguimos? Podemos ser mucho más selectivos para no perder excesivo tiempo analizando todos los significados posibles que a todas luces van a estar incluidos dentro de esos 250 millones de sitios en donde se habla de nuestro vocablo de interés. Ahora nos vamos al scholar.google.com, introducimos la misma palabra inglesa. En menos de 10 segundos contamos con más de 400.000 sitios “técnicos” en donde podemos encontrar información específica sobre el portafolios. Ya hemos ganado mucho en concreción con un mínimo esfuerzo (en vez de ir al buscador general hemos ido al específico). Además, este buscador específico, en función de las citas de que goza cada trabajo y a las cuales nos posibilita el acceso, nos garantiza una primera criba para que no nos den gato por liebre (nos ayuda a no perecer en el intento de manejar y asimilar toda la información de esos 250 millones de sitios virtuales, mucha de la cual es lógicamente redundante, en muchos casos superflua y en otros con significados que poco o nada tienen que ver con el que hemos definido como el nuestro: carpeta o archivo docente).

Pero 400.000 sitios, que a su vez nos remiten a las correspondientes citas, son demasiados para visitarlos. Nos hacemos todavía más selectivos y nos vamos a una base de datos que nos sea familiar. En nuestro caso hablamos de PsycINFO. Queremos saber ahora cuántos artículos de revista, que utilizan la revisión por pares (un nivel de garantía científica más elevado sobre lo que podemos leer), han incluido la palabra *portfolio* en sus títulos. Unos 130. ¡Eureka! Esto ya nos puede parecer más a nuestro alcance y más útil para lograr completar nuestros elementales conocimientos sobre el *portafolios*. No obstante, convendría no olvidar, por un lado, que hemos reducido el ámbito de búsqueda sólo al campo de la psicología y, por otro, que cada uno de estos artículos, en su apartado de Referencias, nos va a remitir a su vez a decenas de otros artículos, capítulos de libros o libros, por lo que de nuevo, no parece que se puedan poner puertas al campo, incluso cuando acotemos al máximo nuestro mundo de preocupación científica: el psicológico. Si en vez de limitarnos a la psicología, entramos en el *sanctasanctorum*, en la *crème de la crème* de la ciencia –la web of science- y pedimos que aparezcan los trabajos (sumamente seleccionados: por revista –índice de impacto- y por contenido –revisión por pares-) con nuestro *portfolio* en el título, nos encontramos nada menos que con más de 4.500. ¿Seguimos? Sólo un poco más. Como hemos hablado de libros, entremos en amazon.com, pinchemos justamente en libros (books) y después escribamos nuestra palabra talismán (*portfolio*). Resultado: 69.849 registros (el 31 de julio de 2006).

Creo no equivocarme si digo que el lector ha captado perfectamente el mensaje irónico de este epílogo metaevaluativo. Ahora sabe a ciencia cierta por qué no escribí un tipo determinado de libro y por qué opté por escribir el ya leído. Con él he pretendido ofrecer un hilo conductor que nos permita ir descubriendo los conceptos más básicos (por lo que se convierten en denominador común del inmenso ámbito de los evaluadores), a la par que específicos (los referidos a la evaluación del profesorado), dejando siempre una multitud de puertas abiertas. A partir de aquí, que cada uno juzgue si, dentro de un análisis de costes/beneficios, le interesa seguir acumulando más conocimientos (ir abriendo o cerrando puertas) a fin de llevar a cabo cuantas especificaciones y matizaciones se juzguen oportunas. Si ésta fuera la decisión (la mía sí lo es, pero con algunos de los matices aquí insinuados), que la suerte y un buen sentido común cultivado nos acompañen, a fin de que no perezcamos abrumados por la abundancia de información y “conocimientos” en, precisamente, una sociedad, como la actual, a la que algunos de nuestros congéneres no se cansan de llamar sociedad del conocimiento y de la información.

Referencias

- Abrami, P.C. (1985). Dimensions of effective college instruction. *The Review of Higher Education*, 8, 211-228.
- Abrami, P.C., D'Apollonia, S., & Rosenfield, S. (1996). The dimensionality of student ratings of instruction: What we know and what we do not know. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. XI, pp. 213-264). New York: Agathon Press.
- Airasian, P.W., & Gullickson, A.R. (1997). *Teacher self-evaluation tool kit*. Thousand Oaks, CA: Corwin Press.
- Aleamoni, L.M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, 13, 153-166.
- Angulo, F., Fernández, J., & Martínez, M. R. (1987). *La evaluación de la enseñanza universitaria*. Madrid: Instituto de Ciencias de la Educación de la Universidad Complutense.
- Ausubel, D.P. (1968/1976). *Psicología Educativa. Un punto de vista cognoscitivo*. México: Trillas
- Bacharach, S.B., Conley, S.C., & Shedd, J.B. (1990/1997). Cómo evaluar al profesorado para promociones y aumentos salariales. En J. Millman y L. Darling-Hammond (Eds.), *Manual para la evaluación del profesorado* (pp. 188-207). Madrid: La Muralla.
- Bailey, G.D. (1981). *Teacher self-assessment: A means for improving classroom instruction*. Washington, DC: National Education Association.
- Barrett, H. (2000). Create your own electronic portfolio. *Learning & Learning with Technology*. <http://electronicportfolios.com/portfolios/iste2k.html> (recuperado el 31/07/06).
- Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment & Evaluation in Higher Education*, 30, 593-601.
- Bergan, J.R. (1982). Contributions of behavioral psychology to school psychology. In T.B. Gutkin & C.R. Reynolds (Eds.), *The handbook of school psychology* (2nd ed., pp. 126-142). New York: Wiley.
- Berk, R.A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research*, 56, 137-172.
- Bickman, L., & Rog, D.J. (1998). *Handbook of applied social research methods*. Thousand Oaks, CA: Sage.
- Bird, T. (1990/1997). El portafolios del profesor: un ensayo sobre las posibilidades. En J. Millman y L. Darling-Hammond (Eds.), *Manual para la evaluación del profesorado* (pp. 332-351). Madrid: La Muralla.
- Black, K., Daiker, D. A., Sommers, J., & Stygall, G. (Eds.).(1994). *New directions in portfolio assessment: Reflective practice, critical theory, and large-scale scoring*. Portsmouth, NH: Boynton.
- Bloom, B.S., Hastings, J.T., & Madaus, G.F. (1971). *Handbook on formative and summative evaluations of students' learning*. New York: McGraw-Hill.
- Bronfenbrenner, V. (1977). Toward a experimental ecology of human development. *American Psychologist*, 32, 513-531.
- Brown, J.D., & Wolfe-Quintero, K. (1997). Teacher portfolios for evaluation: A great idea or a waste of time. *The Language Teacher*, 21, 28-30.
- Cano, E. (2005). *El portafolios del profesorado universitario. Un instrumento para la evaluación y para el desarrollo profesional*. Barcelona: Octaedro.

- Centra, E., & Gaubatz, N. (2000). Is there gender bias in student evaluations of teaching? *Journal of Higher Education*, 71, 17-33.
- Centra, J.A. (1979). *Determining faculty effectiveness*. San Francisco, CA: Jossey-Bass.
- Centra, J.A., & Potter, D.A. (1980). School and teacher effects: An interrelational model. *Review of Educational Research*, 50, 273-291.
- Chelimsky, E., & Shadish, W.R. (Eds.).(1997). *Evaluation for the 21th century: A handbook*. Thousand Oaks, CA: Sage.
- Cohen, E.H. (2005). Student evaluations of course and teacher: Factor analysis and SSA approaches. *Assessment & Evaluation in Higher Education*, 30, 123-136.
- Cook, T. D., & Shadish, W. R. (1986). Program evaluation: The worldly science. *Annual Review of Psychology*, 37, 193-232.
- Cornesky, R., & McCool, S. (1992). Total quality improvement guide for institutions of higher education. Madison, WI: Magna Publications.
- Crispín, M.L y Caudillo, L. (1998). *El uso del portafolios como herramienta para mejorar la calidad de la docencia*. México: Universidad Iberoamericana.
- Darling-Hammond, L., Wise, A.E., & Pease, S.R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, 53, 285-328.
- Davis, J.T., & Swift, L.J. (1995). Teaching portfolios at a research university. *Journal of Excellence in College Teaching*, 6, 101-115.
- Dawoud, M. (1983). Une grille d'évaluation de la compétence du professeur d'université. *Revue de Psychologie Appliquée*, 33, 17-32.
- Duke, D.L., & Stiggins, R.J. (1990/1997). Más allá de la competencia mínima: evaluación para el desarrollo profesional. En J. Millman y L. Darling-Hammond (Eds.), *Manual para la evaluación del profesorado* (pp. 165-187). Madrid: La Muralla.
- Eisner, E. (1979). *The educational imagination*. New York: Macmillan.
- Ellett, C.D., & Teddlie, C. (2003). Teacher evaluation, teacher effectiveness and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education*, 17, 101-128.
- ERS Standards Committee (1983). Evaluation Research Society standards for program evaluation. In American Evaluation Association (Ed.), *New directions for program evaluation* (pp. 7-18). San Francisco: Jossey-Bass.
- Feistritzer, C.E. (1984). *The making of a teacher: A report on teacher education and certification*. Washington, DC: National Center for Educational Information.
- Feldman, K. A. (1976). The superior college teacher from the student's view. *Research in Higher Education*, 5, 243-288.
- Fernández, A. (2004). El portafolios docente como estrategia formativa y de desarrollo profesional. *Educar*, 33, 127-142.
- 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. (1993). Incidencia de las variables de clasificación en la evaluación de la calidad docente. *Gaceta de la Complutense*, 95, 32-35.
- Fernández, J. (1995). Metaevaluación del profesorado de la calidad docente. *Gaceta de la Complutense*, 111, 16-19.
- Fernández, J. (1997). A Spanish evaluation model in higher education: Circular evaluation. *Higher Education Management*, 9, 71-84.
- Fernández, J. (2001). La evaluación dentro del contexto educativo. In J. Fernández (Ed.), *La psicología, una ciencia diversificada*. Madrid: Pirámide.

- Fernández, J., & Mateo, M. A. (1992). Student evaluation of university teaching quality: Analysis of a questionnaire for a sample of university students in Spain. *Educational and Psychological Measurement, 52*, 675-686.
- Fernández, J., & Mateo, M. A. (1993). The development and factorial validation of the Academic Setting Evaluation Questionnaire. *Educational and Psychological Measurement, 53*, 425-435.
- Fernández, J., & Mateo, M. A. (1994). The circular evaluative process of higher education: A Spanish Sample. In OECD (Ed.), *Evaluation and the decision making process in higher education: French, German, and Spanish experiences* (pp. 83-94). Paris: OECD Publications.
- Fernández, J., & Mateo, M. A. (1998). Student and faculty gender in ratings of university teaching. *Sex Roles, 37*, 997-1003.
- Fernández, J., Mateo, M. A., & Muñiz, J. (1995). Evaluation of the academic setting in Spain. *European Journal of Psychological Assessment, 11*, 133-137.
- Fernández, J., Mateo, M. A., & Muñiz, J. (1996). Valoración por parte del profesorado de la evaluación docente realizada por los alumnos. *Psicothema, 8*, 167-172.
- Fitz-Gibbon, C.T., & Morris, L.L. (1987). *How to design a program evaluation*. Newbury Park, CA: Sage.
- Frey, P.W. (1978). A two dimensional analysis of student ratings of instruction. *Research in Higher Education, 9*, 69-91.
- Garbarino, J. (1990). The human ecology at early risk. In S.J. Meisels & J.P. Shonroff (Eds.), *Handbook of early childhood intervention* (pp. 78-96). New York: Cambridge University Press.
- Glasman, N.S., & Biniaminov, I. (1981). Input-ouput analysis of schools. *Review of Educational Research, 51*, 509-539.
- Glass, G.V., & Ellet, F.S. (1980). Evaluation research. *Annual Review of Psychology, 31*, 211-228.
- Good, J.M., & Weaver, A. (2003). Evaluation of preservice teachers' internships: A model to encourage career continuity and program reform. *Journal of Personnel Evaluation in Education, 17*, 263-275.
- Good, T.I., & Mulryan, C. (1990/1997). Clasificación del profesorado: llamada de atención sobre el control y la autoevaluación de profesores. En J. Millman y L. Darling-Hammond (Eds.), *Manual para la evaluación del profesorado* (pp. 265-299). Madrid: La Muralla.
- Gordon, P.A. (1998). *Student evaluations of college instructors: An overview*. <http://chiron.valdosta.edu/whuitt/files/tcheval.pdf> (recuperado el 8/08/06).
- Guba, G.E., & Lincoln, Y.S. (1982). *Effective evaluation*. San Francisco: Jossey-Bass.
- Guba, E.G., & Lincoln, Y.S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Hauge, T.E. (2006). Portfolios and ICT as means of professional learning in teacher education. *Studies in Educational Evaluation, 32*, 23-36.
- Heath, M. (2002). Electronic portfolios for reflective self-assessment. *Teacher Librarian, 30*, 19-23.
- Herman, J.L., Morris, L.L., & Fitz-Gibbon, C.T. (1987). *Evaluator's handbook*. Newbury Park, CA: Sage.
- Hopkins, D. (1989). *Evaluation for school development*. Milton Keynes, PH: Open University Press.
- House, H.R.(1980). *Evaluating with validity*. Beverly Hills, CA: Sage.

- Hoy, A.W., & Spero, R.B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education, 21*, 343-356.
- Joint Committee on Standards for Educational Evaluation –JCSEE-. (1981/1988). *Normas de evaluación para programas, proyectos y material educativo*. México: Trillas.
- Joint Committee on Standards for Educational Evaluation –JCSEE-. (1988). *The personnel evaluation standards: How to assess systems for evaluating educators*. Newbury Park, CA: Sage.
- Joint Committee on Standards for Educational Evaluation –JCSEE-. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage.
- Joint Committee on Standards for Educational Evaluation –JCSEE-. (2003). *The student evaluation standards: How to improve evaluations of students*. Thousand Oaks, CA: Corwin.
- Kellaghan, T., & Stufflebeam, D.L. (Eds.).(2003). *International handbook on educational evaluation*. Dordrecht, NL: Kluwer.
- Kells, H.R. (1995). Creating a culture of evaluation and self-regulation in higher education organizations. *Total Quality Management, 6*, 457-468.
- King, J.A., & Stevahn, L. (2002). Three frameworks for considering evaluator role. In K.E. Ryan & T.A. Schwandt (Eds.), *Exploring evaluator role and identity* (pp. 1-16). Greenwich, CT: Information Age Publishing.
- King, J.A., Stevahn, L., Ghore, G., & Minnema, J. (2001). Toward a taxonomy of essential evaluator competencies. *American Journal of Evaluation, 22*, 229-247.
- Kratz, H.E. (1986). Characteristics of the best teachers as recognized by children. *Pedagogical Seminary, 3*, 413-118.
- Landsheere, G. de (1985). *Diccionario de la evaluación y de la investigación educativas*. Barcelona: Oikos.
- Lester, P.E., & Bishop, L.K. (2000). *Handbook of tests and measurement in education and the social sciences* (2nd ed.). Lanham, MD: The Scarecrow Press.
- Leviton, L.C. (2003). Evaluation use: advances, challenges and applications. *American Journal of Evaluation, 24*, 525-535.
- Lewy, A. (Ed.).(1991). *The international encyclopedia of curriculum*. New York: Pergamon Press.
- Lindzey, G., & Aronson, E. (Eds.).(1985). *Handbook of social psychology* (2nd ed.). New York: Random House.
- Lipsey, M. W., & Cordray, D. S. (2000). Evaluation methods for social intervention. *Annual Review of Psychology, 51*, 345-375.
- MacDonald, B. (1976). Evaluation and the control of education. In D. Tawney (Ed.), *Curriculum evaluation today: Trends and implications* (pp.125-136). London: McMillan.
- Madaus, G.F., Scriven, M.S., & Stufflebeam, D.L. (Eds.).(1983). *Evaluation models*. Boston, MA: Kluwer.
- Marsh, H.W. (1983). Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics. *Journal of Educational Psychology, 75*,150-166.
- Marsh, H.W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*, 707-754.

- Marsh, H.W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388.
- Marsh, H.W., & Roche, L.A. (2000). Effects of grading leniency and low workloads on students' evaluations of teaching: Popular myths, bias, validity or innocent bystanders. *Journal of Educational Psychology*, 92, 202-228.
- Mateo, M.A., & Fernández, J. (1992). Análisis confirmatorio de la estructura dimensional de un cuestionario para la evaluación de la calidad de la enseñanza. *Investigaciones Psicológicas*, 11, 73-82.
- Mateo, M.A., & Fernández, J. (1993). Dimensiones de la calidad de la enseñanza universitaria. *Psicothema*, 5, 265-275.
- Mateo, M.A., & Fernández, J. (1995). Evaluation of the setting in which university faculty carry out their teaching and research functions: The ASEQ. *Educational and Psychological Measurement*, 55, 329-334.
- Mathison, S. (Ed.).(2004). *The encyclopedia of evaluation*. Thousand Oaks, CA: Sage.
- McCormack, C. (2005). Reconceptualizing student evaluation of teaching: An ethical framework for changing times. *Assessment & Evaluation in Higher Education*, 30, 463-476.
- McCornick, R., & James, M. (1983). *Curriculum evaluation in schools*. London: Croom Helm.
- McCutcheon, G. (1982). Qualitative curriculum evaluation. In H.E. Mitzel (Ed.), *Encyclopedia of educational research* (5th. ed., pp. 1503-1505). New York: Free Press.
- McLaughlin, M.W., & Pfeifer, R.S. (1988). *Teacher evaluation: Improvement, accountability and effective learning*. New York: Teachers College Press.
- Medley, D.M. (1972). Early history of research on teacher behaviour. *International Review of Education*, 18, 430-439.
- Meier, R.S., & Feldhusen, J.F. (1979). Another look at Dr. Fox: Effect of stated purpose for evaluation, lecturer expressiveness, and density of lecture content on student ratings. *Journal of Educational Psychology*, 71, 339-345.
- Millman, J. (Ed.).(1981). *Handbook of teacher evaluation*. Beverly Hills, CA: Sage.
- Millman, J., & Darling-Hammond, L. (Eds.).(1990). *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers*. Newbury Park, CA: Sage.
- Ministerio de Educación de Chile (2003). *Manual portafolio. Sistema de evaluación del desempeño profesional docente 2003*. Santiago de Chile: Gobierno de la Nación.
- Miron, M. (1988). Students' evaluation and instructors' self-evaluation of university instruction. *Higher Education*, 17, 175-181.
- Moss, P.M., Schutz, A.M., & Collins, K.M. (1998). An integrative approach to portfolio evaluation for teacher licensure. *Journal of Personnel Evaluation in Education*, 12, 139-161.
- 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.
- Natriello, G. (1990/1997). Consecuencias deseadas e imprevistas: propósitos y efectos de la evaluación del profesorado. En J. Millman y L. Darling-Hammond (Eds.), *Manual para la evaluación del profesorado* (pp. 49-64). Madrid: La Muralla.
- Nevo, D. (1983). The conceptualization of educational evaluation: An analytical review of the literature. *Review of Educational Research*, 53, 117-128.

- Nickerson, R.S., Perkins, D.N., & Smith, E.E. (1985/1987). *Enseñar a pensar*. Barcelona: Paidós/MEC.
- Ory, J.C. (1992). Meta-assessment: Evaluating assessment activities. *Research in Higher Education*, 33, 467-481.
- Parlett, M., & Hamilton, D. (1977). Evaluation as illumination: A new approach to the study of innovative programmes. In D. Hamilton, D. Jenkins, C. King, B. MacDonald & M. Parlett (Eds.), *Beyond the numbers game: A reader in educational evaluation* (pp. 6-22). Basingstoke: Macmillan.
- Patton, M.Q. (1997). *Utilization-focused evaluation: The new century text*. Thousand Oaks, CA: Sage.
- Payne, D.A. (1982). Portrait of the school psychologist as program evaluator. In C.R. Reynolds & T.B. Gutkin (Eds.), *The handbook of school psychology* (pp. 891-915). New York: Wiley.
- Perloff, R., Perloff, E., & Sussna, E. (1976). Program evaluation. *Annual Review of Psychology*, 27, 569-594.
- Peterson, K.D. (1990/1997). Asesoramiento y evaluación para el profesorado principiante. En J. Millman y L. Darling-Hammond (Eds.), *Manual para la evaluación del profesorado* (pp. 147-164). Madrid: La Muralla.
- Perry, R.P., Abrami, P.C., & Leventhal, L. (1979). Educational seduction: The effect of instructor expressiveness and lecture content on student ratings and achievement. *Journal of Educational Psychology*, 71, 107-116.
- Popham, W.J. (1975). *Educational evaluation*. Englewood Cliffs, NJ: Prentice-Hall.
- Pratt, D. (1987). Predicting career success in teaching. *Action in Teacher Education Journal*, 8, 25-34.
- Reason, P. & Bradbury, H. (2001) *Handbook of action research*. Thousand Oaks, CA: Sage.
- Reigeluth, C.M. (Ed.).(1999/2000). *Diseño de la instrucción. Teorías y modelos. Un nuevo paradigma de la teoría de la instrucción*. Madrid: Santillana.
- Rice, J.M. (1897). The futility of the spelling grind. *The Forum*, 23,163-172.
- Rice, J.M. (1914). *Scientific management in education*. New York: Hinds, Noble & Eldredge.
- Rossi, P.H., Freeman, H.E., & Wright, S.R. (1979). *Evaluation: A systematic approach*. Beverly Hills, CA: Sage.
- Rushton, J.P., & Murray, H.G. (1985). On the assessment of teaching effectiveness in British universities. *Bulletin of the British Psychological Society*, 38, 361-365.
- Rutman, L. (1980). *Planning useful evaluations: Evaluability assessment*. Beverly Hills, CA: Sage.
- Rutman, L. (1984). Evaluability assessment. In L. Rutman (Ed.), *Evaluation research methods: A basic guide* (2nd ed., pp.27-38). Newbury Park, CA: Sage.
- Sánchez Matos, E.M. (2001). *Calidad total: organizaciones de calidad, organizaciones de éxito*. Madrid: Libertarias.
- Scheerens, J. (2004). The evaluation culture. *Studies in Educational Evaluation*, 30, 105-124.
- Schunk, D.H. (1991/1997). *Teorías del aprendizaje* (2^{da} ed.). México: Prentice-Hall Hispanoamericana.
- Schunk, D.H., & Zimmerman, B.J. (Eds.).(1994). *Self-regulation of learning and performance: Issues and educational applications*. Hillsdale, NJ: Erlbaum.
- Scriven, M. (1969). An introduction to meta-evaluation. *Educational Product Report*, 2, 36-38.

- Scriven, M. (1974). Pros and cons about goal-free evaluation. *Evaluation Comment*, 3, 1-4.
- Scriven, M. (1990/1997). Selección del profesorado. En J. Millman y L. Darling-Hammond (Eds.), *Manual para la evaluación del profesorado* (pp. 107-146). Madrid: La Muralla.
- Sechrest, L., & Figueredo, A.J. (1993). Program evaluation. *Annual Review of Psychology*, 44, 645-674.
- Shadish, W.R., Jr., & C.S. Reichardt, C.S. (Eds.).(1987). *Evaluation Studies Review Annual* (Vol. 12). Newbury Park, CA: Sage.
- Shores, E., & Grace, C. (2004). *El portafolio paso a paso. Infantil y primaria*. Barcelona: Graó.
- Stake, R.E. (1976a). *Evaluating educational programmes: The need and the response*. Washington, DC: OECD Publications Center.
- Stake, R.E. (1976b). A theoretical statement of responsive evaluation. *Studies in Educational Evaluation*, 2, 19-22.
- Stake, R.E. (1978). The case study method in social inquiry. *Educational Researcher*, 7, 5-8.
- Stecher, B.M., & Davis, W.A. (1987). *How to focus on evaluation*. Newbury Park, CA: Sage.
- Sternberg, R.J., & Spear-Swerling, L. (1996/1999). *Enseñar a pensar*. Madrid: Aula XXI/Santillana.
- Straw, R.B., & Cook, T.D. (1990). Meta-evaluation. In H.J. Walberg & G.D. Haertel (Eds.), *The international encyclopedia of educational evaluation* (pp. 58-60). New York: Pergamon Press.
- Struening, E.L., & Guttentag, M. (1975). *Handbook of evaluation research*. Beverly Hills, CA: Sage.
- Stufflebeam, D.L. (1974). *Meta-evaluation*.
<http://www.wmich.edu/evalctr/pubs/ops/ops03.pdf> (recuperado el 27/07/06).
- Stufflebeam, D.L. (1978). Meta-evaluation: An overview. *Evaluation and the Health Professions*, 1, 17-43.
- Stufflebeam, D.L. (2001). The metaevaluation imperative. *American Journal of Evaluation*, 22, 183-209.
- Stufflebeam, D.L. (2003). The CIPP model for evaluation. In T. Kellaghan & D. Stufflebeam (Eds.), *International handbook of educational evaluation*. Dordrecht, NL: Kluwer.
- Stufflebeam, D.L. (2004). A note on the purposes, development, and applicability of the Joint Committee Evaluation Standards. *American Journal of Evaluation*, 25, 99-102.
- Stufflebeam, D.L., & Shinkfield, A.J.(1985/1987). *Evaluación sistemática. Guía teórica y práctica*. Barcelona: Paidós.
- Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Guba, E.B., Hammond, R.L., Meriman, H.O., & Provus, M. (1971). *Educational evaluation and decision making*. Itasca, Ill: Peacock.
- Suchman, E.A. (1967). *Evaluative research*. New York: Sage.
- Sutherland, S. (2004). Creating a culture of data use for continuous improvement: A case study of an Edison Project School. *American Journal of Evaluation*, 25, 277-293.
- Talmage, H. (1982). Evaluation of programs. In H.E. Mitzel (Ed.), *Encyclopedia of educational research* (5th. ed., pp. 592-611). New York: Free Press.
- Taut, S.M., & Alkin, M.C. (2003). Program staff perceptions of barriers to evaluation implementation. *American Journal of Evaluation*, 24, 213-226.

- Tigelaar, D.E.H., Dolmans, D.H.J.M., Wolfhagen, I.H.A.P., & Van der Vleuten, C.P.M. (2004). Using a conceptual framework and the opinions of portfolio experts to develop a teaching portfolio prototype. *Studies in Educational Evaluation, 30*, 305-321.
- Trevisan, M.S. (2004). Practical training in evaluation: A review of the literature. *American Journal of Education, 25*, 255-272.
- Tsang, M.C. (1988). Cost analysis for educational policymaking: A review of case studies in education in developing countries. *Review of Educational Research, 58*, 181-230.
- Tyler, R.W. (1942). General statement on evaluation. *Journal of Educational Research, 35*, 492-501.
- Tyler, R.W. (1950). *Basic principles of curriculum and instruction*. Chicago, Ill: University of Chicago Press.
- Tyler, R.W. (1967). Changing concepts of educational evaluation. In R.E. Stake (Ed.), *Perspectives of curriculum evaluation* (Vol. 1). New York: Rand McNally.
- Voss, J., Perkins, D.N., & Segal, J.W. (Eds.).(1991). *Informal reasoning and education*. Hillsdale, NJ: Erlbaum.
- Vygotsky, L.S. (1978/1979). *El desarrollo de los procesos psicológicos superiores*. Barcelona: Critica.
- Walberg, H.J., & Haertel, G.D. (Eds.).(1990). *The international encyclopedia of educational evaluation*. New York: Pergamon Press.
- Watkins, D., Marsh, H. W., & Young, D. (1987). Evaluating tertiary teaching: A New Zealand perspective. *Teaching and Teacher Education, 3*, 41-53.
- Watson, J.B. (1926). What the nursery has to say about instincts. In C. Murchison (Ed.), *Psychologies of 1925* (pp. 1-35). Worcester, MA: Clark University Press.
- Weiss, C.H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice, 9*, 5-19.
- Wholey, J.S., Hatry, H.P., & Newcomer, K.E., (Eds.).(1994). *Handbook of Practical Program Evaluation*. San Francisco: Jossey-Bass.
- Wilkerson, D.J., Manatt, R.P., Rogers, M.A., & Maughan, R. (2000). Validation of student, principal, and self-ratings in 360° feedback for teacher evaluation. *Journal of Personnel Evaluation in Education, 14*, 179-192.
- Williams, J.E. (1989). A numerically developed taxonomy of evaluation theory and practice. *Evaluation Review, 13*, 18-31.
- Wilson, T.C. (1988). Student evaluation-of-teaching forms: A critical perspective. *The Review of Higher Education, 12*, 79-95.
- Wise, A., Darling-Hammond, L., McLaughlin, M., & Bernstein, H. (1985). Teacher evaluation: A study of effective practice. *Elementary School Journal, 86*, 61-121.
- Wolf, R.A. (1969). A model for curriculum evaluation. *Psychology in the Schools, 6*, 107-108.
- Wolf, R.L. (1974). The use of judicial evaluation methods in the formulation of educational policy. *Educational Evaluation and Policy Analysis, 1*, 19-28.
- Wortman, P.M. (1983). Evaluation research: A methodological perspective. *Annual Review of Psychology, 34*, 223-260.
- Worthen, B.R., & Sanders, J.R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman.

Anexos

Dimensiones del Profesorado Universitario

En nuestros días, una de las preocupaciones más fundamentales en el terreno de la enseñanza en general, y de la enseñanza universitaria en particular, es el “grado de calidad” de la misma. El corpus bibliográfico actual abunda en señalar aquellas características que parecen tener una incidencia especial en el logro de una buena formación educativa. Con objeto de especificar concretamente esa *calidad educativa*, se presenta a continuación una serie de ítems en relación con la docencia. Su tarea consistirá en valorar, en una escala de 0 a 4, cada uno de los 27 ítems, según la importancia que le atribuya a la hora de conseguir una buena calidad docente. El 0 significa, en todos los casos, nula incidencia (irrelevante), mientras que el 4 implica máxima incidencia (totalmente relevante). El 1, 2 y el 3 significan grados intermedios, de menor a mayor, dentro de los dos extremos valorativos. Las respuestas son, lógicamente, confidenciales. ¿Hay alguna duda?

Muchas gracias por su colaboración

Ítems	Escala
1. Cumplimiento normal de asistencia a clase	0 1 2 3 4
2. Utilidad/valor del trabajo adicional a clase	0 1 2 3 4
3. Presentación clara de exigencias del programa	0 1 2 3 4
4. Las publicaciones del profesor	0 1 2 3 4
5. Los contenidos dados a lo largo del curso	0 1 2 3 4
6. Utilidad/aplicabilidad de los contenidos	0 1 2 3 4
7. El número de años docentes	0 1 2 3 4
8. Dificultad/esfuerzo que exige la asignatura	0 1 2 3 4
9. Tolerancia del profesor hacia distintos enfoques	0 1 2 3 4
10. Accesibilidad del profesor	0 1 2 3 4
11. Ecuanimidad/justicia en las calificaciones	0 1 2 3 4
12. Rango/grado académico del profesor	0 1 2 3 4
13. Utilidad de la bibliografía recomendada	0 1 2 3 4
14. El número de alumnos por grupo	0 1 2 3 4
15. Estimulación de la participación en clase	0 1 2 3 4
16. Coherencia/claridad en las explicaciones	0 1 2 3 4
17. Eficacia en presentación/organización de contenidos	0 1 2 3 4
18. Interés despertado en el alumno por la asignatura	0 1 2 3 4
19. El interés previo del alumno por la materia	0 1 2 3 4
20. La relación entre profesor-alumno	0 1 2 3 4
21. La interacción profesor-grupo	0 1 2 3 4
22. El interés del profesor por la asignatura	0 1 2 3 4
23. El sexo del profesor	0 1 2 3 4
24. El prestigio docente del profesor	0 1 2 3 4
25. El trabajo de investigación del profesor	0 1 2 3 4
26. El nivel de conocimientos del profesor	0 1 2 3 4
27. El nivel de exigencia del profesor	0 1 2 3 4

Instrumento de valoración de los alumnos

El Rectorado de la Universidad Complutense está llevando a cabo un estudio con el objetivo de conocer las opiniones de los estudiantes de esta Universidad sobre la enseñanza que reciben. Con ello se pretende potenciar la **participación real** de cada alumno en la mejora del funcionamiento de la docencia. Se le ruega, pues, asuma esta responsabilidad, leyendo detenidamente el cuestionario que a continuación se le presenta y emitiendo una valoración **sopesada e independiente** sobre cada ítem.

La cumplimentación del cuestionario gozará del más estricto **anonimato**. Sea, por tanto, **sincero** en sus respuestas, procurando mostrar su grado de acuerdo/desacuerdo sobre todos y cada uno de los elementos. Sólo así se podrá obtener una información **válida y fiable** de la realidad educativa de la Universidad Complutense, imprescindible para una intervención eficiente en la mejora de la docencia.

Este trabajo ha sido diseñado y está dirigido por un equipo científico-técnico de nuestra Universidad, aunque la recogida de la información y el análisis de los datos se ha encargado a una empresa ajena al mundo universitario. Los resultados obtenidos gracias a este estudio se darán a conocer una vez comenzado el siguiente curso académico: los globales, a los distintos centros y departamentos y los individuales, a cada profesor.

MUCHAS GRACIAS POR SU COLABORACIÓN

INSTRUCCIONES

Este cuestionario consta de una serie de ítems precedidos por un número. Todos los elementos del cuestionario están redactados en forma afirmativa. Su tarea consistirá en evaluar cada elemento independientemente de los demás, según una escala valorativa de 1 a 7, sabiendo que el 1 siempre indica el **grado más bajo** del contenido sugerido por cada ítem y el 7 el **grado más alto**. Para anotar sus contestaciones, rodee con un círculo el número correspondiente a la calificación que atribuya a cada elemento. Si en alguna ocasión se equivoca en la contestación, tache el número incorrecto y vuelva a rodear el número deseado.

Antes de cumplimentar los ítems de este cuestionario, responda a los datos que aparecen en el apartado de clasificación.

DATOS DE CLASIFICACIÓN

1. Centro:
2. Rama/Sección o especialidad:
3. Asignatura:
4. Profesor:
5. Curso: primero = 1 segundo = 2 tercero = 3 cuarto = 4 quinto = 5
 sexto = 6 adaptación = 7
6. Turno: mañana = 1 tarde = 2 noche = 3
7. Sexo: varón = 1 mujer = 2
8. Edad:
9. Repite asignatura: sí = 1 no = 2
10. Asignatura: anual = 1 primer cuatrimestre = 2 segundo cuatrimestre = 3
11. Asistencia a clase: 0-24% = 1 25-49% = 2 50-74% = 3 75-100% = 4
12. Cuestionario cumplimentado: primera vez = 1 más veces = 2

RECUERDE LAS INSTRUCCIONES DE CUMPLIMENTACIÓN DE ESTE CUESTIONARIO: EL 1 INDICA EL **GRADO MÁS BAJO** DE ACUERDO CON CADA AFIRMACIÓN, MIENTRAS QUE EL 7 EL **GRADO MÁS ALTO**.

1. Su forma de impartir la clase me ha ayudado a comprender mejor la materia explicada	1 2 3 4 5 6 7
2. Muestra interés por las cuestiones planteadas por cada alumno	1 2 3 4 5 6 7
3. Motiva a hacer preguntas sobre cualquier duda o problema en torno a lo expuesto	1 2 3 4 5 6 7
4. Despierta interés en el alumno por la asignatura	1 2 3 4 5 6 7
5. Se muestra accesible en su relación con los alumnos	1 2 3 4 5 6 7
6. Su actitud facilita en el alumno un sentimiento de libertad para expresar su propias ideas	1 2 3 4 5 6 7
7. En sus explicaciones alude a las implicaciones derivadas de los distintos enfoques	1 2 3 4 5 6 7
8. Establece una adecuada interrelación entre los distintos temas de la asignatura	1 2 3 4 5 6 7
9. En sus explicaciones ha quedado reflejada la utilidad de los contenidos expuestos	1 2 3 4 5 6 7
10. Manifiesta entusiasmo por los contenidos explicados	1 2 3 4 5 6 7
11. Como resultado de lo que he aprendido en esta asignatura, ahora me siento más interesado por esta materia	1 2 3 4 5 6 7
12. Responde a las preguntas del alumno con precisión	1 2 3 4 5 6 7
13. La forma de explicación del profesor consiguió mantener mi atención durante las clases	1 2 3 4 5 6 7
14. Explica de forma coherente los contenidos de cada tema	1 2 3 4 5 6 7
15. Se esfuerza por atraer la atención del alumno hacia la materia impartida	1 2 3 4 5 6 7
16. Explica con claridad	1 2 3 4 5 6 7
17. Estimula la participación de los alumnos en la clase	1 2 3 4 5 6 7
18. Se muestra competente en los contenidos que imparte	1 2 3 4 5 6 7
19. Manifiesta una preparación previa de la materia explicada en clase	1 2 3 4 5 6 7
20. Es constante en la asistencia a clase	1 2 3 4 5 6 7
21. Lo explicado en la clase se ajusta a los objetivos expresados en el programa de la asignatura	1 2 3 4 5 6 7
22. El procedimiento de examen que utilizado por el profesor permite reflejar adecuadamente los conocimientos del alumno	1 2 3 4 5 6 7
23. Examina preferentemente sobre los contenidos considerados como fundamentales en el programa	1 2 3 4 5 6 7
24. Sus calificaciones en los exámenes responden a los niveles de conocimiento demostrados por los alumnos	1 2 3 4 5 6 7
25. La bibliografía básica recomendada en el programa sirve para que los alumnos puedan preparar adecuadamente la materia	1 2 3 4 5 6 7
26. Me gustaría recibir la enseñanza de otra asignatura con este profesor	
DESDE UNA CONSIDERACIÓN GENERAL, EVALÚE GLOBALMENTE (LA ESCALA DE VALORACIÓN SIGUE SIENDO DE 1 A 7):	
27. La eficacia docente de ese este profesor	1 2 3 4 5 6 7
CALIFIQUE COMPARATIVAMENTE (EN LA MISMA ESCALA DE 1 A 7)	
28. La docencia de este profesor con respecto a los restantes profesores del curso actual	1 2 3 4 5 6 7
29. A esta asignatura con respecto a las restantes del curso actual, según la contribución que haya supuesto para su formación académica	1 2 3 4 5 6 7

Instrumento de valoración de los profesores

El Rectorado está interesado en conocer su opinión en torno a una serie de cuestiones que afectan muy directamente la marcha de la Universidad Complutense. Sus respuestas a este cuestionario ofrecerán una información válida sobre algunos aspectos importantes del funcionamiento de nuestra Universidad, que contribuirá, sin duda, a mejorar la situación actual.

La recogida de la información se realiza por una Empresa ajena al ámbito universitario. El tratamiento estadístico global de los datos garantiza el **absoluto anonimato** de las respuestas. Se le ruega la mayor **sinceridad** en las mismas, ya que sólo así se podrá obtener un diagnóstico fiable de la Universidad Complutense.

Muchas gracias por su valiosa colaboración

En la página siguiente se le presentan una serie de elementos redactados en forma afirmativa, seguidos de una escala valorativa creciente que va del **1** al **7**, siendo así que el 1 indica siempre el **grado más bajo** del contenido de cada elemento y el 7 el **grado más alto**.

Su tarea consistirá en rodear con un círculo el número correspondiente a su opinión en torno a cada elemento. Le rogamos encarecidamente que cumplimente todos y cada uno de los ítems.

1. Las condiciones materiales en que realizo mi trabajo son satisfactorias	1 2 3 4 5 6 7
2. Se me ofrecen posibilidades financieras para llevar a cabo mi investigación	1 2 3 4 5 6 7
3. Se me dan facilidades institucionales para la publicación de mis trabajos	1 2 3 4 5 6 7
4. Considero justamente remuneradas mis actividades como profesor universitario	1 2 3 4 5 6 7
5. Me siento respaldado por los compañeros en las actividades que realizo como profesor	1 2 3 4 5 6 7
6. El contexto académico fomenta mi labor profesional	1 2 3 4 5 6 7
7. Existe una adecuada comunicación académica entre los miembros de mi departamento	1 2 3 4 5 6 7
8. Las relaciones humanas con los compañeros de mi departamento favorecen mi actividad académica-profesional	1 2 3 4 5 6 7
9. Los sistemas de selección del profesorado son los adecuados	1 2 3 4 5 6 7
10. Los sistemas de control de la actividad docente del profesorado son correctos	1 2 3 4 5 6 7
11. Existen criterios claros para la evaluación de las actividades investigadoras	1 2 3 4 5 6 7
12. Existe concordancia entre mis expectativas de lo que debiera ser un profesor y lo que de hecho es	1 2 3 4 5 6 7
13. La sociedad aprecia la labor realizada por el profesor universitario	1 2 3 4 5 6 7
14. Las instituciones universitarias estimulan mi perfeccionamiento como docente	1 2 3 4 5 6 7
15. Los sistemas de promoción del profesorado son correctos	1 2 3 4 5 6 7
16. Se me ha preparado institucionalmente para llevar a cabo de forma adecuada mis funciones como investigador	1 2 3 4 5 6 7
17. Existen perspectivas favorables para realizar mi trabajo como docente	1 2 3 4 5 6 7
18. Las perspectivas futuras como investigador universitario son favorables	1 2 3 4 5 6 7
19. Encuentro facilidades institucionales para resolver los problemas profesionales como profesor universitario	1 2 3 4 5 6 7
20. Cuento con el tiempo adecuado para poder realizar mis tareas de investigación	1 2 3 4 5 6 7
21. Los alumnos manifiestan interés por la materia que imparto	1 2 3 4 5 6 7
22. Las instituciones universitarias fomentan mi actividad investigadora	1 2 3 4 5 6 7
23. El sistema de funcionariado es el correcto para desempeñar las funciones docentes y de investigación del profesorado	1 2 3 4 5 6 7
24. Los alumnos consultan sus dudas/problemas en el horario destinado a “recepción de alumnos”	1 2 3 4 5 6 7
25. Se tienen en cuenta las opiniones de los alumnos sobre la enseñanza impartida a fin de mejorar la propia docencia	1 2 3 4 5 6 7
26. Los profesores del departamento colaboran en la elaboración/realización de los programas de investigación del mismo	1 2 3 4 5 6 7
27. Cada año tomo en consideración las opiniones de los alumnos a la hora de establecer la correspondiente metodología docente	1 2 3 4 5 6 7
28. Acomodo mi docencia a las peculiaridades/exigencias de cada grupo de alumnos	1 2 3 4 5 6 7
29. El trabajo de los alumnos se ajusta sin dificultad a las demandas de mi asignatura	1 2 3 4 5 6 7
30. El sistema de contratos laborales posibilitaría un mejor desarrollo de las funciones del profesor	1 2 3 4 5 6 7
31. Es conveniente algún sistema de evaluación “objetivo” para la labor investigadora de cada profesor	1 2 3 4 5 6 7

32. Los alumnos muestran una valoración diferencial de cada profesor en función de su calidad docente	1 2 3 4 5 6 7
33. Me siento ayudado en mis investigaciones por los compañeros del departamento	1 2 3 4 5 6 7
EN GENERAL:	
34. Me hallo satisfecho con las condiciones que me brindan las instituciones universitarias para ejercer mis funciones como profesor	1 2 3 4 5 6 7
35. Encuentro un clima intradepartamental propicio para desempeñar mis funciones de profesor	1 2 3 4 5 6 7
36. Los alumnos participan activamente en el funcionamiento de la docencia	1 2 3 4 5 6 7
37. Los sistemas actuales de selección y promoción del profesorado son los adecuados	1 2 3 4 5 6 7
FINALMENTE, UTILIZANDO LA MISMA ESCALA VALORATIVA DE 1 A 7, RESPONDA TAMBIÉN A LOS SIGUIENTES ELEMENTOS:	
38. Los alumnos se cuestionan la posible utilidad de los contenidos impartidos	1 2 3 4 5 6 7
39. La evaluación de los profesores realizada por los alumnos proporciona información útil a la institución sobre su calidad de enseñanza	1 2 3 4 5 6 7
40. El número de alumnos asignado es el correcto para prepararlos profesionalmente	1 2 3 4 5 6 7
41. Los alumnos proponen formas alternativas al modo de dar la clase	1 2 3 4 5 6 7
42. Existe colaboración intradepartamental en la programación docente	1 2 3 4 5 6 7
43. Enseño la asignatura para la que realmente me encuentro más capacitado	1 2 3 4 5 6 7
44. Puedo enfocar la asignatura desde la perspectiva que juzgo más conveniente	1 2 3 4 5 6 7
45. La formación recibida para ejercer como profesor es suficiente	1 2 3 4 5 6 7

FUNCIONAMIENTO DE LOS SERVICIOS INSTITUCIONALES

Ahora, por favor, evalúe, con el mismo formato de escala, el funcionamiento de los siguientes servicios institucionales, sabiendo que el 1 indica un **pésimo funcionamiento** y el 7 un **funcionamiento óptimo**. En el caso de que no exista alguno de los servicios en su centro o no conozca su funcionamiento, rodee con un círculo el “N/E-N/S”, que corresponde respectivamente a “no existe o no sabe”.

1. Biblioteca	1 2 3 4 5 6 7 N/E-N/S
2. Laboratorios	1 2 3 4 5 6 7 N/E-N/S
3. Hemeroteca	1 2 3 4 5 6 7 N/E-N/S
4. Medios informáticos	1 2 3 4 5 6 7 N/E-N/S
5. Medios audiovisuales	1 2 3 4 5 6 7 N/E-N/S
6. Secretaría	1 2 3 4 5 6 7 N/E-N/S
7. Condiciones físicas del inmueble	1 2 3 4 5 6 7 N/E-N/S
8. Actividades culturales organizadas por la institución	1 2 3 4 5 6 7 N/E-N/S
9. Personal auxiliar (Bedeles...)	1 2 3 4 5 6 7 N/E-N/S
10. Decanato	1 2 3 4 5 6 7 N/E-N/S
11. Rectorado	1 2 3 4 5 6 7 N/E-N/S
12. Instalaciones deportivas	1 2 3 4 5 6 7 N/E-N/S
13. Comedores y cafeterías	1 2 3 4 5 6 7 N/E-N/S
14. Ministerio de Educación y Ciencia	1 2 3 4 5 6 7 N/E-N/S

DATOS DEMOGRÁFICOS Y DEL CURRÍCULUM ACADÉMICO

Por favor, escriba sobre la raya o rodee los números correspondientes en su caso. Ponga una X en las casillas que juzgue adecuadas a su situación en el ítem número 10.

1. Edad -----
2. Sexo: varón = 1 mujer = 2
3. Centro docente -----
4. Departamento -----
5. Sección Departamental -----
6. Categoría: numerario = 1 no numerario = 2
7. Turno en el que imparte docencia: mañana = 1 tarde = 2 noche = 3
8. Curso: primero = 1 segundo = 2 tercero = 3 cuarto = 4 quinto = 5
 sexto = 6 adaptación = 7
9. Antigüedad en el centro: de 1 a 5 años = 1 de 6 a 10 años = 2 de 11 a 15 años = 3 de 16 a 20 años = 4 de 21 a 25 años = 5 de 26 a 30 años = 6
 de 31 a 35 años = 7 de 36 a 40 años = 8 de 41 a 45 años = 9
10. Estime el tiempo promedio aproximado que en los últimos 5 años académicos ha venido dedicando a cada una de las siguientes actividades:

	0-19%	20-39%	40-59%	60-79%	80-100%
Investigación					
Docencia					
Servicios					

MUCHAS GRACIAS POR SU COLABORACIÓN

Devolución de la información en el modelo de evaluación circular

La evaluación de los alumnos

GUÍA PARA LA INTERPRETACIÓN DE LOS RESULTADOS

Estimado Profesor/a:

A fin de facilitar la interpretación de los resultados referidos al cuestionario de evaluación de la calidad docente, para el curso académico 1990-91, se formulan las siguientes *indicaciones* relativas a las hojas de datos.

En la página 4 aparece la “FICHA TÉCNICA”, donde se presentan algunos de los datos más relevantes de este trabajo. Conviene resaltar dos aspectos fundamentales:

- a) Los alumnos que respondieron al cuestionario son los que habitualmente asistían a clase en ese momento (bien al final del primer cuatrimestre, o bien al final del curso), lo que garantiza una “validez ecológica” a las conclusiones derivadas de la presente investigación, tanto a escala individual como global.
- b) El cuestionario con el que se ha recogido la información garantiza satisfactoriamente la fiabilidad y validez de la misma (véanse los índices de fiabilidad y validez en la *ficha técnica*, incluida en la página cuatro).

La página 5 contiene la información que le permitirá conocer las *opiniones* de los alumnos sobre la enseñanza recibida. En el recuadro superior se proporciona el número de los alumnos que evaluaron la docencia y algunas de sus características de tipo demográfico (edad y sexo). En la parte izquierda se incluyen todos los elementos del cuestionario, redactados en su forma abreviada. A la derecha de los mismos aparecen algunos estadísticos de tendencia central y de dispersión. Se presentan la *media* (valor alrededor del cual se distribuyen las puntuaciones), la *mediana* (puntuación que deja por encima y por debajo de sí el 50% de los casos) y la *moda* (puntuación más repetida).

Cualquiera de los tres índices permite la aplicación del siguiente **PRINCIPIO GENERAL**: *las puntuaciones por encima de 4 (punto medio de la escala valorativa) ponen de manifiesto una consideración positiva, mientras que las puntuaciones inferiores a este punto medio han de interpretarse en la dirección de una evaluación más bien negativa.*

De esta forma es posible obtener una valoración de la propia actividad docente, siempre en relación con las *opiniones de los alumnos* que cumplimentaron el cuestionario.

Se ofrece, igualmente, la *desviación típica* (dispersión de las puntuaciones) y la *distribución de frecuencias* (cuántos alumnos le puntuaron en cada uno de los siete valores de la escala) a fin de que complete y matice el juicio llevado a cabo a través del análisis de los estadísticos de tendencia central.

A título de ejemplo, una media de 4 obtenida por dos profesores puede tener una significación y valoración muy distinta. Así ocurre si a uno de ellos todos los alumnos le valoran en torno a 4 (una desviación típica muy pequeña), mientras que a otro la mitad le asigna un 1 y la otra mitad le otorga un 7 (una desviación típica considerablemente superior).

En esta misma hoja, en la parte derecha, se incluye una representación gráfica de la media obtenida en cada uno de los 29 elementos, de forma que a simple vista se puedan captar los *puntos más fuertes y los más débiles* de la evaluación que han realizado los alumnos sobre la docencia.

Los datos de la segunda gráfica (página 6) posibilitan una valoración centrada en siete NÚCLEOS TEÓRICOS y un bloque de elementos ponderativos/criterio, de probada relevancia a la hora de comprender la calidad docente de un profesor. También aquí se aplicará el **principio general de valoración** para cada uno de los núcleos: *si es superior a 4, la evaluación de la docencia por parte de los alumnos es positiva, pero si es inferior ha de ser considerada más bien negativa.*

Hasta aquí la valoración se ha referido exclusivamente al análisis de las puntuaciones, tomando como término de comparación la *media teórica* (igual a 4).

Esta evaluación podría resultar incompleta y, por ende, no todo lo precisa que sería deseable. Para evitar esto, se ofrecen más datos en la página 7, con los que se pretende matizar y enriquecer la información, ya que permiten establecer una *evaluación comparativa* con el entorno inmediato en el que se desempeña la función docente. Con los resultados que se presentan en esta gráfica se puede ir comparando la evaluación individual con las evaluaciones de la Asignatura, del Departamento (o, en su caso, de la Sección Departamental), de la Facultad o Escuela (o Rama o Sección), del tipo de Centro (Experimental, No Experimental, Escuela Universitaria) y, por fin, de la Universidad.

En el caso de la Asignatura o Sección Departamental sólo se ofrecerán datos cuando sea posible salvaguardar la confidencialidad, es decir, cuando haya 3 ó más profesores evaluados. Para las comparaciones con el Departamento y el Centro se ha seguido siempre esta *regla*: los resultados presentados se refieren a la “unidad” de rango inferior, es decir, Sección Departamental frente a Departamento y Rama/Sección frente a Centro, cuando se dé esta posibilidad. Para las comparaciones entre la evaluación del Profesor, por una parte, y la de la Asignatura, de la Sección Departamental o del Departamento, por otra, se ha procedido de la forma más *rigurosa y matizada* posible, es decir, sustrayendo la valoración del Profesor de la total, correspondiente a cada una de las

unidades antes aludidas. En el resto de comparaciones posibles esta puntuación no ha sido excluida, dada su presumiblemente escasa incidencia.

En la octava página se le ofrece la posibilidad de constatar gráficamente las posibles diferencias entre las valoraciones referidas al Profesor, al Departamento (o Sección Departamental, siempre y cuando se cuente con tres o más profesores evaluados) y al Centro (Rama o Sección).

En la parte superior de la hoja aparecen las dos primeras evaluaciones, en función de los dos factores del cuestionario. El primero comprende los 4 núcleos teóricos iniciales de los 4 representados en la gráfica de la página 6. El segundo se corresponde con el quinto núcleo, de idéntica denominación. En ambos casos, las barras representan las puntuaciones factoriales (estimación estadística de la puntuación del sujeto –o, en su caso, “unidad”- en el factor) correspondientes a los 3 elementos de comparación: Profesor, Departamento y Centro. Las dos siguientes comparaciones están fundamentadas en el sexto y séptimo núcleos teóricos, referidos a los “requisitos docentes” y a la “ecuanimidad evaluadora”. La última comparación tiene como base uno de los criterios generales más significativos: la eficacia docente de cada profesor.

En las páginas 9 y 10 aparecen las gráficas comparativas de los dos últimos cursos académicos. En la página 11, se presentan juntas las puntuaciones de los cursos 1989-90 y 1990-91. Gracias a la información de estas tres últimas páginas es posible establecer fácilmente una *comparación interanual* (como es

lógico sólo se recibirá esta información si ha sido posible valorar la docencia en los referidos cursos).

De este modo se puede constatar la *estabilidad, el empeoramiento o la mejora* de la enseñanza impartida durante este tiempo y sacar las conclusiones oportunas.

Es necesario subrayar que este procedimiento de evaluación de la docencia es sólo *uno de los múltiples posibles*, aunque ciertamente el más extendido y el que ofrece una mayor garantía. Lo que se obtiene a través del mismo es *el conocimiento de las opiniones de los alumnos en torno a la enseñanza que reciben*.

A la luz de estos resultados, bien por sí mismo o bien, si así lo prefiere, a través de las consultas oportunas al Equipo investigador que ha diseñado y dirigido este estudio, tiene la oportunidad de *optimizar la función docente*.

FICHA TÉCNICA

- CENTROS EVALUADOS: Todos los de la Universidad, salvo las recientemente creadas Escuela Superior de Informática y la Escuela Universitaria de Biblioteconomía y Documentación. Tampoco se realizó, en esta ocasión, la recogida de información en el Hospital Clínico.
- PROFESORES EVALUADOS: a) Facultades: 3.353 b) Escuelas Universitarias: 488 c) TOTAL: 3.841
- ASIGNATURAS EVALUADAS: a) Facultades: 1.500 b) Escuelas Universitarias: 398 c) TOTAL: 1.898
- UNIDADES PROFESOR/ASIGNATURA: a) Facultades: 4.186 b) Escuelas Universitarias: 776 c) TOTAL: 4.962
- ALUMNOS EVALUADORES: a) Facultades: 61.966 b) Escuelas Universitarias: 9.605 c) TOTAL: 71.571
- EVALUACIONES REALIZADAS: a) Facultades: 204.221 b) Escuelas Universitarias: 32.641 c) TOTAL: 236.862
- INSTRUMENTO DE EVALUACIÓN: Cuestionario de Evaluación del Profesorado Universitario de la Complutense.
- PRINCIPALES PROPIEDADES PSICOMÉTRICAS DEL CUESTIONARIO:
 - a) VALIDEZ TEÓRICA: todos los análisis factoriales (PCA) llevados a cabo (uno con la “muestra” total y 41 con diferentes “submuestras” significativas) mostraron invariablemente los dos factores previstos *a priori*. La varianza explicada por estos dos factores giró en torno al 97%. La correlación entre ambos estuvo en torno a .77.
 - b) FIABILIDAD: un alpha de .96 en la muestra general y un rango que va desde .95 a .98 en las 41 submuestras.
- La recogida y análisis de los datos se realizó por una Empresa totalmente ajena al ámbito universitario: AREA, Gabinete de Estudios y Aplicaciones, S.A. La recogida de la información se llevó a cabo entre el 15 de enero y el 6 de febrero para las asignaturas del primer cuatrimestre y entre el 24 de marzo y el 22 de mayo para las asignaturas del segundo cuatrimestre y las del curso completo.

Profesor: APELLIDOS Y NOMBRE DEL PROFESOR

Asignatura: NOMBRE DE LA ASIGNATURA

Centro: NOMBRE DEL CENTRO

NUCLEOS TEORICOS	ELEMENTOS DEL CUESTIONARIO	MEDIA	NIVEL EN EL NUCLEO (promedio de las medias)						
			1	2	3	4	5	6	7
ESTILO DOCENTE	1. Su forma de enseñar facilita la comprensión de la materia	5.7							
	4. Despierta interés en el alumno por la asignatura	5.7							
	10. Manifiesta entusiasmo por los contenidos explicados	6.2							
	13. Su explicación mantuvo mi atención en las clases	5.7							
	15. Atrae la atención del alumno en la materia explicada	5.5							
ESTRUCTURACION DE CONTENIDOS	7. Alude a las implicaciones de los distintos enfoques	5.4							
	8. Establece una adecuada interrelación entre los temas	5.8							
	9. Refleja la utilidad de los contenidos expuestos	5.8							
CLARIDAD EXPOSITIVA	12. Responde a las preguntas con precisión	5.4							
	14. Explica de forma coherente los contenidos de cada tema	5.8							
	16. Explica con claridad	5.6							
COMPETENCIA ACADEMICA	18. Se muestra competente en los contenidos que imparte	6.3							
	19. Manifiesta una adecuada preparación previa a la clase	6.4							
HABILIDADES MOTIVACIONALES Y DE INTERACCION	2. Muestra interés por las inquietudes de cada alumno	5.4							
	3. Motiva a preguntar sobre cualquier duda	4.8							
	5. Se muestra accesible en su relación con los alumnos	5.0							
	6. Facilita la expresión de las ideas propias de los alumnos	4.9							
	17. Estimula la participación del alumno en la clase	4.7							
REQUISITOS DOCENTES	20. Es constante en la asistencia a clase	6.7							
	21. Lo explicado se ajusta a los objetivos del programa	5.9							
	25. La bibliografía recomendada sirve para preparar la materia	5.0							
ECUANIMIDAD EVALUADORA	22. El procedimiento de examen permite al alumno reflejar sus conocimientos	5.1							
	23. Examina sobre contenidos considerados como fundamentales	5.7							
	24. Las notas reflejan el nivel de conocimientos mostrados por los alumnos	4.9							
CRITERIOS GENERALES	11. Gracias a lo aprendido siento mayor interés por esta área	5.8							
	26. Me gustaría recibir de nuevo otra asignatura con este profesor	5.4							
	27. La eficacia docente de este profesor	6.0							
	28. Calificación comparativa del profesor con los restantes del curso	5.9							
	29. Calificación comparativa de esta asignatura con las restantes del curso	6.0							

GRAFICO DE MUESTRA. LOS DATOS NO CORRESPONDEN A NINGUNA UNIDAD PROFESOR-ASIGNATURA REAL. IGUAL OCURRE CON EL RESTO DE LAS UNIDADES DE ANALISIS

Profesor: APELLIDOS Y NOMBRE DEL PROFESOR

Asignatura: NOMBRE DE LA ASIGNATURA

Centro: NOMBRE DEL CENTRO

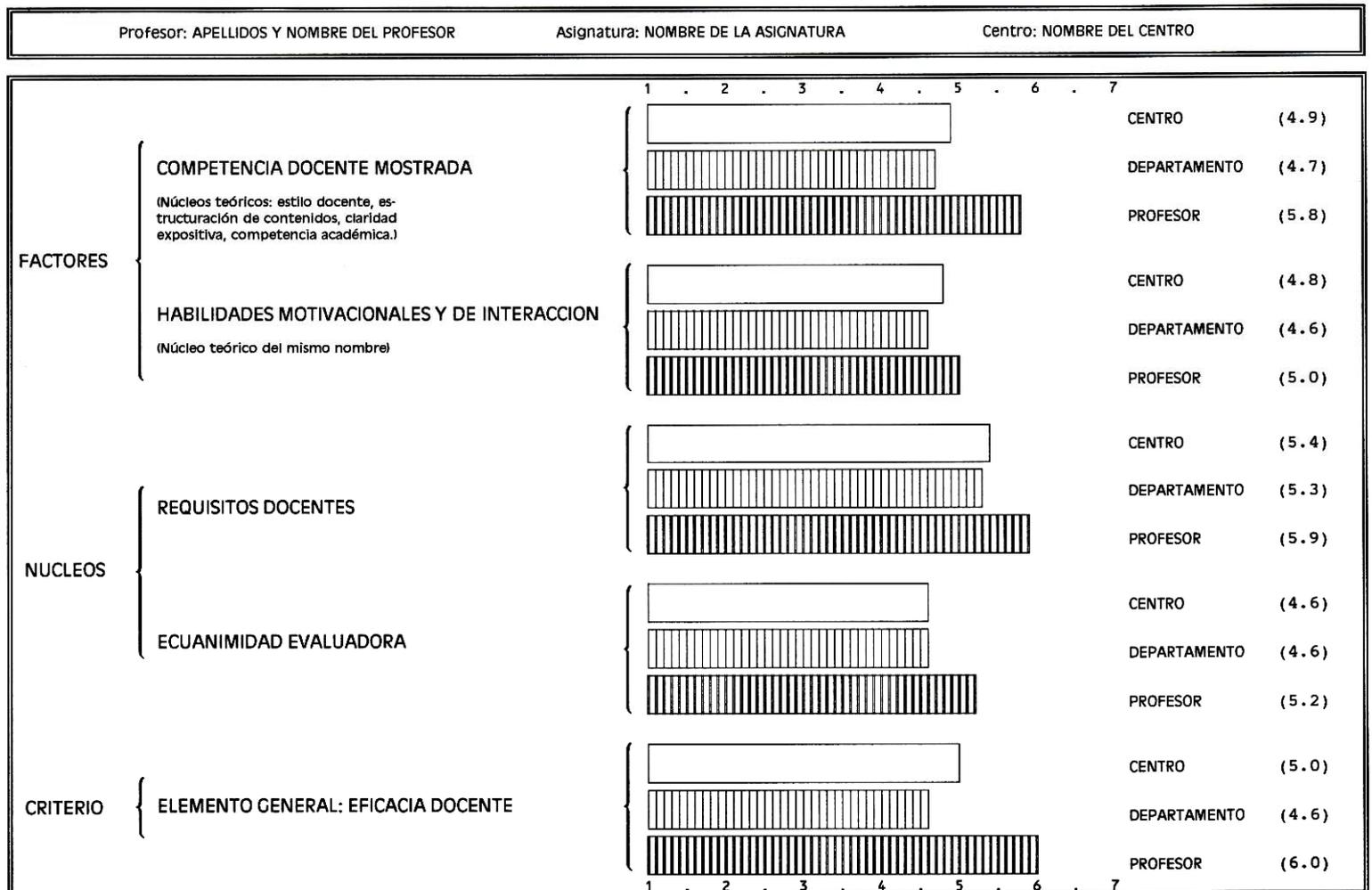
ELEMENTOS DEL CUESTIONARIO	PROFESOR		ASIGNATURA		DPTO. (1)		CENTRO (2)		TIPO CENTRO (3)		UNIVERSIDAD	
	ME DIA	DESVI. TIPICA	ME DIA	DESVI. TIPICA	ME DIA	DESVI. TIPICA	ME DIA	DESVI. TIPICA	ME DIA	DESVI. TIPICA	ME DIA	DESVI. TIPICA
1. Su forma de enseñar facilita la comprensión de la materia	5.7	1.2	4.5	2.1	4.6	1.9	4.9	1.7	4.4	2.0	4.3	1.8
2. Muestra interés por las inquietudes de cada alumno	5.4	1.2	4.8	1.9	4.9	1.8	5.1	1.6	5.0	1.7	5.2	1.9
3. Motiva a preguntar sobre cualquier duda	4.8	1.5	4.4	1.9	4.6	1.8	4.6	1.8	4.3	1.8	4.6	1.8
4. Despierta interés en el alumno por la asignatura	5.7	1.3	4.5	2.1	4.6	1.9	4.4	1.8	4.7	1.6	4.2	1.8
5. Se muestra accesible en su relación con los alumnos	5.0	1.4	4.6	2.1	4.7	2.1	5.0	1.4	5.2	1.8	5.1	1.7
6. Facilita la expresión de las ideas propias de los alumnos	4.9	1.5	4.4	2.0	4.5	1.9	4.3	1.7	4.9	1.8	4.7	1.8
7. Alude a las implicaciones de los distintos enfoques	5.4	1.2	4.8	1.8	4.8	1.7	5.2	1.6	4.3	1.7	4.4	1.6
8. Establece una adecuada interrelación entre los temas	5.8	1.1	4.6	1.9	4.7	1.8	5.1	1.6	4.9	1.7	4.5	1.5
9. Refleja la utilidad de los contenidos expuestos	5.8	1.2	4.6	2.0	4.8	1.9	4.5	1.8	4.7	1.3	4.6	1.8
10. Manifiesta entusiasmo por los contenidos explicados	6.2	1.1	4.7	1.9	5.0	1.8	5.2	1.6	5.2	1.7	5.1	1.8
11. Gracias a lo aprendido siento mayor interés por esta área	5.8	1.3	4.6	2.0	4.8	1.9	4.1	1.9	4.4	2.0	4.1	2.1
12. Responde a las preguntas con precisión	5.4	1.3	4.4	2.1	4.6	1.9	4.7	1.8	4.6	1.8	4.7	1.1
13. Su explicación mantuvo mi atención en las clases	5.7	1.3	4.2	2.3	4.4	2.1	4.2	1.9	4.1	2.1	4.3	2.3
14. Explica de forma coherente los contenidos de cada tema	5.8	1.2	4.5	2.2	4.6	2.0	5.7	1.7	4.8	1.8	4.7	1.8
15. Atrae la atención del alumno en la materia explicada	5.5	1.2	4.5	2.0	4.6	1.8	4.3	1.5	4.3	1.8	4.8	1.6
16. Explica con claridad	5.6	1.3	4.4	2.2	4.5	2.1	4.7	1.8	4.6	1.5	4.3	2.3
17. Estimula la participación del alumno en la clase	4.7	1.4	4.2	1.9	4.2	1.7	4.9	1.7	4.4	1.8	4.3	1.5
18. Se muestra competente en los contenidos que imparte	6.3	0.9	5.1	1.9	5.2	1.8	5.3	1.2	5.2	1.7	5.1	1.4
19. Manifiesta una adecuada preparación previa a la clase	6.4	0.8	4.9	2.0	5.1	1.9	5.8	1.5	5.5	1.8	5.4	1.3
20. Es constante en la asistencia a clase	6.7	0.6	6.1	1.3	5.8	1.6	6.1	1.4	5.9	1.6	6.3	1.5
21. Lo explicado se ajusta a los objetivos del programa	5.9	1.2	5.2	1.6	5.3	1.6	5.4	1.9	5.3	1.7	5.2	1.7
22. El procedimiento de examen permite al alumno reflejar sus conocimientos ..	5.1	1.4	4.9	1.2	4.7	1.5	4.6	1.5	4.8	1.9	4.8	1.6
23. Examina sobre contenidos considerados como fundamentales	5.7	1.2	4.8	1.2	4.8	1.4	4.9	1.7	4.8	1.5	4.1	1.7
24. Las notas reflejan el nivel de conocimientos mostrados por los alumnos	4.9	1.5	4.6	1.4	4.4	1.5	4.2	1.3	4.3	1.9	4.6	1.9
25. La bibliografía recomendada sirve para preparar la materia	5.0	1.7	4.6	1.7	4.8	1.7	4.6	1.8	4.4	1.9	4.1	1.8
26. Me gustaría recibir de nuevo otra asignatura con este profesor	5.4	1.8	4.2	2.2	4.3	2.1	4.4	2.7	4.1	2.1	4.0	2.1
27. La eficacia docente de este profesor	6.0	1.1	4.5	2.0	4.6	1.9	5.0	1.6	4.6	1.7	4.7	1.8
28. Calificación comparativa del profesor con los restantes del curso	5.9	1.1	4.5	2.1	4.6	2.0	4.8	1.7	4.2	1.4	4.9	1.7
29. Calificación comparativa de esta asignatura con las restantes del curso	6.0	1.1	5.4	1.7	5.4	1.6	4.3	1.6	4.6	1.8	4.7	1.7

(1) Departamento o Sección Departamental.

(2) Facultad (o Rama/Sección), Escuela Universitaria.

(3) Experimental, No experimental, Escuela Universitaria

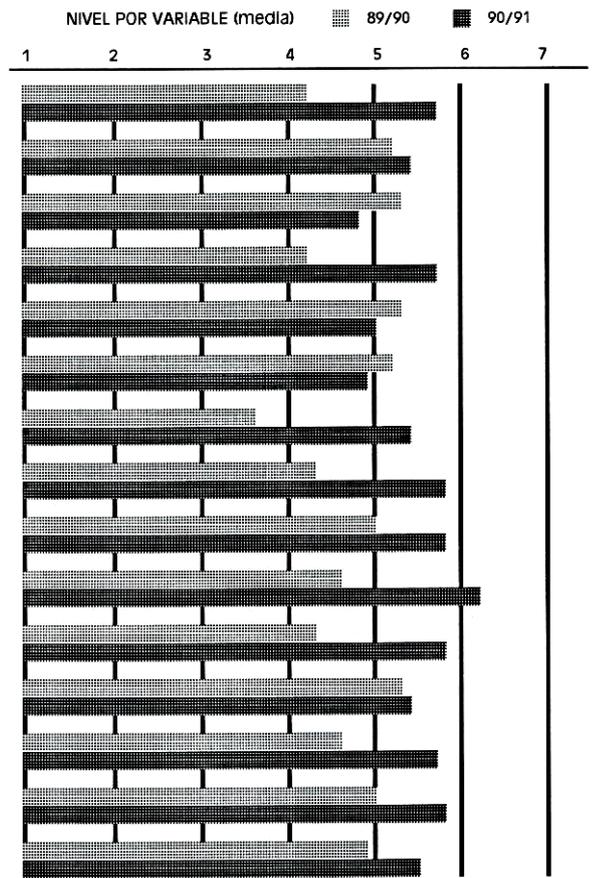
GRAFICO DE MUESTRA. LOS DATOS NO CORRESPONDEN A NINGUNA UNIDAD PROFESOR-ASIGNATURA REAL. IGUAL OCURRE CON EL RESTO DE LAS UNIDADES DE ANALISIS



*Las barras correspondientes a los "FACTORES" representan la puntuación factorial obtenida en cada uno de ellos, las de los "NUCLEOS" el promedio de las medias en los mismos, y las del "CRITERIO" la media en el elemento n°27.
 GRAFICO DE MUESTRA. LOS DATOS NO CORRESPONDEN A NINGUNA UNIDAD PROFESOR-ASIGNATURA REAL. IGUAL OCURRE CON EL RESTO DE LAS UNIDADES DE ANALISIS

Profesor: APELLIDOS Y NOMBRE DEL PROFESOR	Asignatura: NOMBRE DE LA ASIGNATURA	Centro: NOMBRE DEL CENTRO
---	-------------------------------------	---------------------------

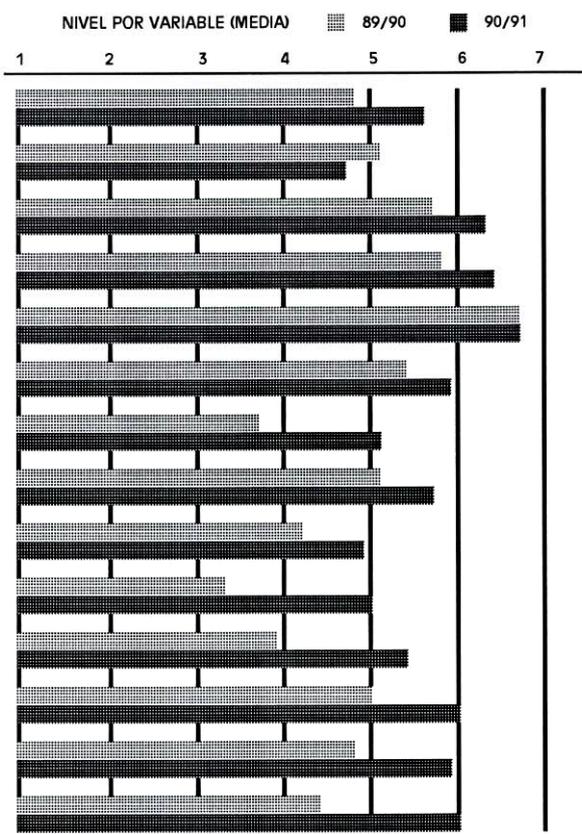
ELEMENTOS UNO A QUINCE DEL CUESTIONARIO	MEDIA	
	89/90	90/91
1. Su forma de enseñar facilita la comprensión de la materia	4.2	5.7
2. Muestra Interés por las inquietudes de cada alumno	5.2	5.4
3. Motiva a preguntar sobre cualquier duda	5.3	4.8
4. Despierta Interés en el alumno por la asignatura	4.2	5.7
5. Se muestra accesible en su relación con los alumnos	5.3	5.0
6. Facilita la expresión de las ideas propias de los alumnos	5.2	4.9
7. Alude a las implicaciones de los distintos enfoques	3.6	5.4
8. Establece una adecuada interrelación entre los temas	4.3	5.8
9. Refleja la utilidad de los contenidos expuestos	5.0	5.8
10. Manifiesta entusiasmo por los contenidos explicados	4.6	6.2
11. Gracias a lo aprendido siento mayor interés por esta área	4.3	5.8
12. Responde a las preguntas con precisión	5.3	5.4
13. Su explicación mantuvo mi atención en las clases	4.6	5.7
14. Explica de forma coherente los contenidos de cada tema	5.0	5.8
15. Atrae la atención del alumno en la materia explicada	4.9	5.5



****GRAFICO DE MUESTRA. LOS DATOS NO CORRESPONDEN A NINGUNA UNIDAD PROFESOR-ASIGNATURA REAL. IGUAL OCURRE CON EL RESTO DE LAS UNIDADES DE ANALISIS****

Profesor: APELLIDOS Y NOMBRE DEL PROFESOR	Asignatura: NOMBRE DE LA ASIGNATURA	Centro: NOMBRE DEL CENTRO
---	-------------------------------------	---------------------------

ELEMENTOS DIECISEIS A VEINTINUEVE DEL CUESTIONARIO	MEDIA	
	89/90	90/91
16. Explica con claridad	4.8	5.6
17. Estimula la participación del alumno en la clase	5.1	4.7
18. Se muestra competente en los contenidos que imparte	5.7	6.3
19. Manifiesta una adecuada preparación previa a la clase	5.8	6.4
20. Es constante en la asistencia a clase	6.7	6.7
21. Lo explicado se ajusta a los objetivos del programa	5.4	5.9
22. El procedimiento de examen permite al alumno reflejar sus conocimientos	3.7	5.1
23. Examina sobre contenidos considerados como fundamentales	5.1	5.7
24. Las notas reflejan el nivel de conocimientos mostrados por los alumnos ..	4.2	4.9
25. La bibliografía recomendada sirve para preparar la materia	3.3	5.0
26. Me gustaría recibir de nuevo otra asignatura con este profesor	3.9	5.4
27. La eficacia docente de este profesor	5.0	6.0
28. Calificación comparativa del profesor con los restantes del curso	4.8	5.9
29. Calificación comparativa de esta asignatura con las restantes del curso	4.4	6.0



****GRAFICO DE MUESTRA. LOS DATOS NO CORRESPONDEN A NINGUNA UNIDAD PROFESOR-ASIGNATURA REAL. IGUAL OCURRE CON EL RESTO DE LAS UNIDADES DE ANALISIS****

La evaluación de los profesores

DATOS PARA UNA VALORACION POR UNIDAD

Unidad: NOMBRE DEL CENTRO	Nº total de profesores = 333	Edad Media =41.2	Desv.Típica = 0.3	Varones = 170	Mujeres =152	N.C.Sexo =11
---------------------------	------------------------------	------------------	-------------------	---------------	--------------	--------------

ELEMENTOS DEL CUESTIONARIO (1)	ME DIA	ME DIA	MO DA NA (2)	DES. TIPI.	DISTRIBUCION DE FRECUENCIAS (3)							NIVEL POR VARIABLE (MEDIA)									
					1	2	3	4	5	6	7	1	2	3	4	5	6	7			
A) SATISFACCION/INSATISFACCION ANTE LAS INSTITUCIONES UNIVERSITARIAS																					
1. Las condiciones materiales en que realizo mi trabajo son satisfactorias	3.2	3.0	3	1.5	41	82	86	60	37	22	5										
2. Se me ofrecen posibilidades financieras para realizar mi investigación	2.6	2.5	2	1.3	73	89	77	56	18	7	2										
3. Se me dan facilidades institucionales para publicar mis trabajos	2.6	2.2	1	1.6	105	70	58	30	30	14	6										
4. Considero justamente remuneradas mis actividades como profesor	2.9	2.8	3	1.4	52	82	89	60	29	11	4										
11. Existen criterios claros para evaluar las actividades investigadoras	2.9	2.8	3	1.3	49	85	91	61	29	6	2										
12. Concordancia entre expectativas y realidad de ser profesor	2.4	2.3	2	1.2	79	101	77	46	10	2	1										
13. La sociedad aprecia la labor realizada por el profesor universitario	3.2	3.1	3	1.4	36	73	96	67	42	14	4										
14. Las instituciones universitarias estimulan mi perfeccionamiento	3.0	2.9	2	1.4	48	87	80	58	34	21	2										
16. Preparación institucional para ejercer mis funciones de investigador	2.8	2.6	2	1.5	73	86	72	58	23	13	6										
17. Existen perspectivas favorables para realizar mi trabajo como docente	3.4	3.3	3	1.4	28	65	96	71	46	20	5										
18. Las perspectivas futuras como investigador universitario son favorables	2.4	2.1	1	1.4	107	92	70	28	18	7	5										
19. Facilidades institucionales para la resolución de mis problemas	2.8	2.5	1	1.6	85	81	66	44	25	24	6										
20. Cuento con el tiempo adecuado para realizar mis tareas de investigación	3.1	2.8	2	1.6	61	85	54	56	41	19	10										
22. Las instituciones universitarias fomentan mi actividad investigadora	2.2	1.8	1	1.4	141	80	41	35	20	6	3										
34. Satisfacción general ante las condiciones institucionales	3.4	3.3	3	1.4	29	67	84	62	67	18	5										

(1) Redacción abreviada de los elementos del cuestionario.

(2) En aquellos casos en que no exista una sola moda aparecerá en esta columna un guión.

(3) La suma de frecuencias de cada elemento no siempre tiene que coincidir con el número total de profesores.

LOS DATOS NO CORRESPONDEN A NINGUNA UNIDAD DE ANALISIS REAL

DATOS PARA UNA VALORACION POR UNIDAD

Unidad: NOMBRE DEL CENTRO	Nº total de profesores = 106	Edad Media =41.5	Desv.Típica = 8.8	Varones = 38	Mujeres = 66	N.C.Sexo = 2
---------------------------	------------------------------	------------------	-------------------	--------------	--------------	--------------

ELEMENTOS DEL CUESTIONARIO (1)	ME DÍA	ME DÍA	MO DA NA (2)	DESV. TÍPI.	DISTRIBUCION DE FRECUENCIAS (3)							NIVEL POR VARIABLE (MEDIA)						
					1	2	3	4	5	6	7	1	2	3	4	5	6	7
B) CLIMA SOCIAL PROFESORAL																		
5. Me siento respaldado por los compañeros en las actividades que realizo	3.2	2.9	3	1.7	15	27	29	13	9	7	6							
6. El contexto académico fomenta mi labor profesional	5.1	5.7	7	1.9	8	5	10	11	14	23	35							
7. Existe adecuada comunicación académica entre los miembros de mi dpto.	4.5	4.8	6	1.9	12	9	10	14	20	21	18							
8. Las relaciones con los compañeros de mi dpto. favorecen mi actividad ac.	5.6	5.8	6	1.2	0	3	3	12	23	36	29							
26. Colaboración intradepartamental en realización de programas de investig.	4.8	5.2	6	1.8	7	7	10	17	16	28	20							
33. Me siento apoyado en mis investigaciones por los compañeros del dpto.	4.7	5.0	5	1.9	8	11	8	13	26	18	20							
35. Clima intradepartamental	4.4	4.6	4	1.8	10	12	7	22	21	18	15							
C) RELACIONES CON LOS ALUMNOS																		
21. Los alumnos manifiestan interés por la materia que imparto	4.2	4.5	6	1.9	9	16	12	16	19	21	12							
24. Los alumnos consultan sus dudas en el horario destinado a recepción al.	4.7	4.9	5	1.5	3	8	8	22	33	22	10							
25. Se tienen en cuenta las opiniones de los alumnos sobre la docencia	5.1	5.5	6	1.7	4	7	9	14	19	27	26							
27. Considero las opiniones de los alumnos para estab.la metodología docente	5.2	5.6	6	1.6	2	8	6	11	24	29	26							
28. Acomodo mi docencia a las peculiaridades de cada grupo de alumnos	4.0	4.2	5	2.1	17	17	11	12	18	16	15							
29. El trabajo de los alumnos se ajusta a las demandas de mi asignatura	4.5	4.6	5	1.5	4	6	15	24	29	19	8							
32. Valoración diferencial por el alumnado de la calidad docente	4.9	5.1	6	1.6	3	5	13	16	22	25	18							
36. Participación activa de los alumnos	3.9	3.9	-	1.7	12	12	21	21	20	11	8							

(1) Redacción abreviada de los elementos del cuestionario.

(2) En aquellos casos en que no exista una sola moda aparecerá en esta columna un guión.

(3) La suma de frecuencias de cada elemento no siempre tiene que coincidir con el número total de profesores.

LOS DATOS NO CORRESPONDEN A NINGUNA UNIDAD DE ANALISIS REAL

DATOS PARA UNA VALORACION POR UNIDAD

Unidad: NOMBRE DEL CENTRO	Edad Media =41.2	Desv.Típica = 0.3	Varones = 170	Mujeres =152	N.C.Sexo =11
Nº total de profesores = 333					

ELEMENTOS DEL CUESTIONARIO (1)	ME DIA	MO DA NA (2)	DES. TIPI.	DISTRIBUCION DE FRECUENCIAS (3)							NIVEL POR VARIABLE (MEDIA)							
				1	2	3	4	5	6	7	1	2	3	4	5	6	7	
D) SELECCION Y PROMOCION																		
9. Los sistemas de selección del profesorado son los adecuados	2.7	2.5	2	1.4	80	84	77	54	18	13	3							
10. Los sistemas de control de la actividad docente son correctos	5.2	5.5	6	1.5	7	16	18	50	70	91	73							
15. Los sistemas de promoción del profesorado son correctos	3.1	3.0	3	1.6	66	57	79	64	33	24	7							
23. El sistema de funcionariado es el correcto para ser profesor	2.8	2.3	1	1.8	106	74	50	37	26	20	19							
30. Los contratos laborables posibilitarian mejorar las funciones del prof.	3.6	3.6	4	1.8	49	52	47	70	43	32	21							
31. Conveniencia de sistema"objetivo"para la valoración de la investigación	2.6	2.4	2	1.3	80	89	77	53	20	6	2							
37. Los sistemas actuales de selección y promoción son adecuados	2.9	2.7	1	1.6	80	69	74	55	31	16	7							
E) ELEMENTOS PONDERATIVOS																		
38. Los alumnos se cuestionan la utilidad de los contenidos impartidos	2.7	2.3	1	1.6	92	91	59	41	26	13	7							
39. La evaluación docente de los alumnos es útil a la institución	6.0	6.2	7	1.2	1	8	7	23	39	119	136							
40. El número de alumnos es el correcto para prepararles profesionalmente	5.9	6.3	7	1.4	4	6	13	29	43	91	143							
41. Los alumnos proponen formas alternativas al modo de dar la clase	4.3	4.5	5	1.6	18	42	44	61	77	69	18							
42. Existe colaboración intradepartamental en la programación docente	5.0	5.4	6	1.7	9	26	37	41	62	89	68							
43. Enseño la asignatura para la que realmente me encuentro más capacitado	3.3	3.2	4	1.7	58	63	62	68	35	28	12							
44. Puedo enfocar la asignatura desde la perspectiva que juzgo conveniente	4.3	4.6	6	1.8	29	40	41	50	64	65	40							
45. La formación recibida para ejercer como profesor es suficiente	3.0	2.6	1	1.8	88	71	47	52	36	20	15							

(1) Redacción abreviada de los elementos del cuestionario.

(2) En aquellos casos en que no exista una sola moda aparecerá en esta columna un guión.

(3) La suma de frecuencias de cada elemento no siempre tiene que coincidir con el número total de profesores.

** LOS DATOS NO CORRESPONDEN A NINGUNA UNIDAD DE ANALISIS REAL **

DATOS PARA UNA VALORACION POR UNIDAD

Unidad: NOMBRE DEL CENTRO	Nº total de profesores = 333	Edad Media =41.2	Desv.Típica = 0.3	Varones = 170	Mujeres =152	N.C.Sexo =11
---------------------------	------------------------------	------------------	-------------------	---------------	--------------	--------------

ELEMENTOS DEL CUESTIONARIO	ME DIA	ME DIA	MO DA NA (1)	DES. TIPI.	DISTRIBUCION DE FRECUENCIAS (2)							NIVEL POR VARIABLE (MEDIA)											
					1	2	3	4	5	6	7	1	2	3	4	5	6	7					
F) FUNCIONAMIENTO DE LOS SERVICIOS INSTITUCIONALES																							
1. Funcionamiento de la biblioteca	3.0	3.0	3	1.5	53	50	66	55	33	10	4												
2. Funcionamiento de los laboratorios	4.9	5.2	6	1.6	8	22	36	46	71	92	44												
3. Funcionamiento de la hemeroteca	4.3	4.7	5	1.9	16	18	16	16	29	28	20												
4. Funcionamiento de los medios informáticos	3.3	3.4	1	1.9	21	9	8	13	11	10	3												
5. Funcionamiento de los medios audiovisuales	4.1	4.5	5	1.9	26	27	28	26	47	39	20												
6. Funcionamiento de la secretaría	5.1	5.5	6	1.5	11	14	19	50	68	108	49												
7. Condiciones físicas del inmueble	4.1	4.2	5	1.7	21	31	44	38	57	38	19												
8. Actividades culturales organizadas por la institución	2.9	2.6	1	1.7	69	60	40	42	30	18	4												
9. Funcionamiento del personal auxiliar (bedeles...)	4.2	4.4	6	1.9	34	46	46	41	56	64	36												
10. Funcionamiento del decanato	4.5	4.8	5	1.8	20	28	22	42	63	53	34												
11. Funcionamiento del rectorado	4.3	4.4	4	1.6	19	27	47	60	59	55	22												
12. Funcionamiento de las instalaciones deportivas	3.4	3.3	1	1.9	43	34	22	33	29	22	8												
13. Funcionamiento de los comedores y cafeterías	3.5	3.5	4	1.8	54	49	46	56	49	32	12												
14. Funcionamiento del Ministerio de Educación y Ciencia	3.0	2.8	2	1.7	68	79	60	49	39	21	10												

(1) En aquellos casos en que no exista una sola moda aparecerá en esta columna un guión.

(2) La suma de frecuencias de cada elemento no siempre tiene que coincidir con el número de total de profesores.

LOS DATOS NO CORRESPONDEN A NINGUNA UNIDAD DE ANALISIS REAL

Instrumento de autoevaluación del profesorado

Evaluación de la calidad docente por parte del profesorado

Hasta ahora cada profesor ha recibido **única y exclusivamente los resultados** de la evaluación de la docencia, llevada a cabo a través de las **opiniones de sus alumnos**. Ahora se pretende abrir una nueva vía **complementaria** de la anterior: **la evaluación procedente del propio profesor**. A tal efecto, se le ofrece a continuación un cuestionario formado por elementos que hacen referencia a la forma y al modo de impartir su **docencia**. Trate de cumplimentarlo con total sinceridad, dado que el **principal beneficiario será usted mismo**, al serle devueltos estos resultados junto con los de sus alumnos.

Aunque a todas luces parece conveniente e incluso necesario el ofrecer esta oportunidad al profesor para que opine sobre los mismos aspectos de los que está siendo objeto de evaluación por parte de sus alumnos, sin embargo, es oportuno llamar aquí la atención sobre el bien conocido **sesgo de la deseabilidad social** (“soy extraordinario en todo lo que hace referencia a mi docencia”). A fin de no tener que descartar esta valiosa fuente de información, trate de evitar a toda costa este sesgo.

Muchas gracias por su colaboración

Instrucciones

Este cuestionario consta de una serie de elementos precedidos por un número y redactados en forma afirmativa. Su tarea consistirá en valorar cada elemento independientemente de los demás, según una escala valorativa de 1 a 7, sabiendo que el 1 siempre significa el **grado más bajo** del contenido sugerido por cada elemento y el 7 el **grado más alto**. Para anotar sus contestaciones, rodee con un círculo el número correspondiente a la calificación que atribuya a cada elemento. Si en alguna ocasión se equivoca en la contestación, tache el número incorrecto y vuelva a rodear el número deseado.

1. Mi forma de impartir la clase ha ayudado a los alumnos a comprender mejor la materia explicada	1 2 3 4 5 6 7
2. Muestro interés por las cuestiones planteadas por cada alumno	1 2 3 4 5 6 7
3. Motivo a hacer preguntas sobre cualquier duda o problema en torno a lo expuesto	1 2 3 4 5 6 7
4. Despierto interés en el alumno por la asignatura	1 2 3 4 5 6 7
5. Me muestro accesible en mi relación con los alumnos	1 2 3 4 5 6 7
6. Mi actitud facilita en el alumno un sentimiento de libertad para expresar su propias ideas	1 2 3 4 5 6 7
7. En mis explicaciones aludo a las implicaciones derivadas de los distintos enfoques	1 2 3 4 5 6 7
8. Establezco una adecuada interrelación entre los distintos temas de la asignatura	1 2 3 4 5 6 7
9. En mis explicaciones ha quedado reflejada la utilidad de los contenidos expuestos	1 2 3 4 5 6 7
10. Manifiesto entusiasmo por los contenidos explicados	1 2 3 4 5 6 7
11. Después de haber cursado esta asignatura, mis alumnos parecen sentirse más interesados por esta materia	1 2 3 4 5 6 7
12. Respondo a las preguntas del alumno con precisión	1 2 3 4 5 6 7
13. Mi forma de explicar consiguió mantener la atención de los alumnos durante las clases	1 2 3 4 5 6 7
14. Explico de forma coherente los contenidos de cada tema	1 2 3 4 5 6 7
15. Me esfuerzo por atraer la atención del alumno hacia la materia impartida	1 2 3 4 5 6 7
16. Explico con claridad	1 2 3 4 5 6 7
17. Estimulo la participación de los alumnos en la clase	1 2 3 4 5 6 7
18. Me muestro competente en los contenidos que imparto	1 2 3 4 5 6 7
19. Manifiesto una preparación previa de la materia explicada en clase	1 2 3 4 5 6 7
20. Soy constante en la asistencia a clase	1 2 3 4 5 6 7
21. Lo explicado en la clase se ajusta a los objetivos expresados en el programa de la asignatura	1 2 3 4 5 6 7

22. El procedimiento de examen que utilizo permite reflejar adecuadamente los conocimientos del alumno	1 2 3 4 5 6 7
23. Examino preferentemente sobre los contenidos considerados como fundamentales en el programa	1 2 3 4 5 6 7
24. Las calificaciones en los exámenes responden a los niveles de conocimiento demostrados por los alumnos	1 2 3 4 5 6 7
25. La bibliografía básica recomendada en el programa sirve para que los alumnos puedan preparar adecuadamente la materia	1 2 3 4 5 6 7
26. A mis alumnos les gustaría tenerme de nuevo como profesor	1 2 3 4 5 6 7
27. Me considero un docente eficaz	1 2 3 4 5 6 7

Índice de autores

- Abrami, P.C., 22, 57, 58
Airasian, P.W., 54
Aleamoni, L.M., 22
Alkin, M.C., 26
Angulo, F., 57, 62
Aronson, E., 22
Ausubel, D.P., 16
- Bacharach, S.B., 52
Bailey, G.D., 54
Barrett, H., 56
Beran, T., 88
Bergan, J.R., 14
Berk, R.A., 52
Bernstein, H., 54
Bickman, L., 22
Biniaminov, I., 37
Bird, T., 55
Bishop, L.K., 62
Black, K., 56
Bloom, B.S., 21
Bradbury, H., 22
Bronfenbrenner, V., 17
Brown, J.D., 56
- Cano, E., 55
Caudillo, L., 56
Centra, E., 88
Centra, J.A., 57, 101
Chelimsky, E., 21
Cohen, E.H., 88
Collins, K.M., 56
Conley, S.C., 52
Cook, T.D., 22, 41
Cordray, D.S., 22
Cornesky, R., 38
Crispín, M.L., 56
- Daiker, D.A., 52
D'Apollonia, S., 22
Darling-Hammond, L., 21, 22, 53, 54, 59
Davis, J.T., 56
Davis, W.A., 23, 35
Dawoud, M., 65
Dolmans, D.H.J.M., 55
Duke, D.L., 51
- Eisner, E., 39
- Ellet, F.S., 22, 31
Ellett, C.D., 101
ERS, Standards Committee, 23, 43
- Feistritzer, C.E., 49
Feldhusen, J.F., 58
Feldman, K.A., 57, 58
Fernández, A., 56
Fernández, J., 13, 19, 29, 35, 57, 62, 63, 68, 69, 73, 74, 85, 86, 87, 88, 90
Figueredo, A.J., 22
Fitz-Gibbon, C.T., 21, 22
Foley, W.J., 40
Freeman, H.E., 37
Frey, P.W., 57
- Garbarino, J., 17
Gaubatz, N., 88
Gephart, W.J., 40
Ghere, G., 24
Glasman, N.S., 37
Glass, G. V., 22, 31
Good, J. M., 49
Good, T.I., 53
Gordon, P.A., 22
Grace, C., 56
Guba, G.E., 20, 35, 40
Gullickson, A.R., 54
Guttentag, M., 21
- Haertel, G.D., 21
Hamilton, D., 35
Hammond, R.L., 40
Hastings, J.T., 21
Hatry, H.P., 21
Hauge, T.E., 56
Heath, M., 56
Herman, J.L., 21
Hopkins, D., 102
House, H.R., 35, 36, 39
Hoy, A.W., 49
- James, M., 102
Joint Committee on Standards for Educational Evaluation, JCSEE, 23, 43, 85
- Kellaghan, T., 21

Kells, H.R., 27
 King, J.A., 24
 Kratz, H.E., 53

Landsheere, G. de, 22
 Lester, P.E., 62
 Leventhal, L., 58
 Leviton, L.C., 95
 Lewy, A., 16, 102
 Lincoln, Y.S., 20, 35
 Lindzey, G., 22
 Lipsey, M.W., 22

MacDonald, B., 35
 Madaus, G.F., 19, 21
 Manatt, R.P., 99
 Marsh, H.W., 22, 56, 57, 58, 65, 67, 69, 84, 88
 Martínez, M. R., 57, 62
 Mateo, M. A., 57, 62, 68, 73, 74, 85, 86, 90
 Mathison, S., 21
 Maughan, R., 99
 McCool, S., 38
 McCormack, C., 104
 McCormick, R., 102
 McCutcheon, G., 35
 McLaughlin, M.W., 25, 54
 Medley, D.M., 53
 Meier, R.S., 58
 Merriman, H.O., 40
 Millman, J., 21, 46, 53, 59
 Ministerio de Educación de Chile, 56
 Minnema, J., 24
 Miron, M., 65
 Morris, L.L., 21, 22
 Moss, P.M., 56
 Mulryan, C., 53
 Muñoz, J., 74, 85, 86
 Murray, H.G., 65, 85

Natriello, G., 47
 Nevo, D., 22
 Newcomer, K.E., 21
 Nickerson, R.S., 26

Ory, J.C., 41

Parlett, M., 35
 Patton, M.Q., 94

Payne, D.A., 40, 41
 Pease, S.R., 22
 Perkins, D.N., 26, 34
 Perloff, E., 22
 Perloff, R., 22
 Perry, R.P., 58
 Peterson, K.D., 50
 Pfeifer, R.S., 25
 Popham, W.J., 35, 38
 Potter, D.A., 101
 Pratt, D., 48
 Provus, M., 40

Reason, P., 22
 Reichardt, C.S., 21
 Reigeluth, C.M., 16
 Rice, J.M., 19, 20
 Roche, L.A., 88
 Rog, D.J., 22
 Rogers, M.A., 99
 Rosenfield, S., 22
 Rossi, P.H., 37
 Rushton, J.P., 65
 Rutman, L., 29, 30

Sánchez Matos, E.M., 38
 Sanders, J.R., 21, 43
 Scheerens, J., 27
 Schunk, D.H., 14, 16
 Schutz, A.M., 56
 Scriven, M., 19, 39, 41, 50
 Sechrest, L., 22
 Segal, J.W., 34
 Shadish, W. R., 21, 22
 Shadish, W.R., Jr., 21
 Shedd, J.B., 52
 Shinkfield, A.J., 20
 Shores, E., 56
 Smith, E.E., 26
 Sommers, J., 56
 Spear-Swerling, L., 8
 Spero, R.B., 49
 Stake, R.E., 35, 39
 Stecher, B.M., 23, 35
 Sternberg, R.J., 8
 Stevahn, L., 24
 Stiggins, R.J., 51
 Straw, R.B., 41
 Struening, E.L., 21

Stufflebeam, D.L., 19, 20, 21, 32, 38, 40,
41, 43, 85
Stygall, G., 56
Suchman, E.A., 38
Sutherland, S., 39
Sussna, E., 22
Swift, L.J., 56

Talmage, H., 35
Taut, S.M., 26
Teddlie, C., 101
Tigelaar, D.E.H., 55
Trevisan, M.S., 22
Tsang, M.C., 37
Tyler, R.W., 19, 20, 38

Van der Vleuten, C.P.M., 55
Violato, C., 88
Voss, J., 34
Vygotsky, L.S., 17

Walberg, H.J., 21
Watkins, D., 65
Watson, J.B., 14
Weaver, A., 49
Weiss, C.H., 94
Wholey, J.S., 21
Wilkerson, D.J., 99
Williams, J.E., 35
Wilson, T.C., 58
Wise, A., 22, 54
Wolf, R.A., 35
Wolf, R.L., 39
Wolfe-Quintero, K., 56
Wolfhagen, I.H.A.P., 55
Wortman, P.M., 22
Worthen, B.R., 21, 43
Wright, S.R., 37

Young, D., 65

Zimmerman, B.J., 16