

EVALUACION DE LA CONDUCTA DOCENTE Y DEL CLIMA PSICO-SOCIAL DE CLASE

Luis Miguel VILLAR ANGULO

Departamento de Didáctica y
Organización Escolar.
Universidad de Sevilla.

Marco conceptual de la enseñanza y la evaluación

La Ley de Reforma Universitaria es el marco jurídico que ha propiciado la discusión de los educadores en torno al tópico de la evaluación de la enseñanza universitaria. Las experiencias evaluativas existentes en nuestro país se han multiplicado conforme los estatutos de las distintas universidades han entrado en vigor. Así, la razón, el *porqué* de la evaluación universitaria, hay que encontrarlo en un pronunciamiento político-social: exigir responsabilidad del profesor universitario. Esta razón es la que ha impulsado a los educadores a buscar los procedimientos –el *cómo* evaluar–, sin que hayan participado ni ellos ni los profesores en la justificación de la evaluación.

Aceptado el hecho que la evaluación de la enseñanza universitaria es una práctica ya iniciada en nuestras universidades, con lo cual nos adherimos a una corriente educativa que tiene ya su historia en otros países, prioritariamente anglosajones (ver Elton, 1987), es notorio observar que esta evaluación se está ejerciendo en la enseñanza superior, mientras que los otros niveles -E.G.B. y Enseñanzas Medias- permanecen al margen de este tipo de control, cuando hay países, Estados Unidos, por ejemplo, que le prestan gran atención (véase el número monográfico de *Educational Leadership*, 1987). La evaluación de la enseñanza a este respecto se ha hecho siguiendo las pautas de la investigación sobre la enseñanza, razón por la cual algunos autores, como McGreal, han señalado que se deben establecer ciertas precauciones sobre esta tendencia evaluativa que, de otra parte, es la predominante (Brandt, 1987b).

La revisión de la investigación sobre la enseñanza superior realizada por Dunkin y Barnes (1986) ha seguido un esquema que trata de dar respuesta a tres tipos de preguntas: «¿Cómo se comportan los profesores?», «¿Por qué se comportan como lo hacen?» y «¿Cuáles son los efectos de su conducta?» que dan pie a series de relaciones entre variables. No es nuestra intención aquí analizar su revisión, sino más bien enfatizar cómo la investigación sobre la enseñanza superior se ha restringido a varias áreas: métodos de enseñanza, facetas de la conducta docente, y evaluación y mejora de la enseñanza. La investigación sobre facetas (cognitiva, socioemocional, substantiva y de comunicación) muestra cómo en la enseñanza superior se ha venido aplicando un paradigma de enseñan-

za que ha entrado en desuso en los niveles no universitarios. Por ejemplo, y dentro de la faceta cognitiva, la taxonomía de Bloom ha sido el proceso más investigado en escuelas universitarias durante los años ochenta. De la misma forma, para averiguar cualidades socioemocionales del proceso didáctico se han usado preferentemente las categorías de Flanders. Los sistemas observacionales han servido para recopilar frecuencias de conductas docentes y posteriormente relacionarlas con medidas de producto (rendimiento, satisfacción, absentismo, etc. del estudiante). No obstante, la investigación en enseñanza superior ha dado lugar a dimensiones comportamentales específicas de nivel, como la claridad en la enseñanza y el entusiasmo o la expresividad que se incluyó en numerosos inventarios de evaluación. Como Dunkin y Barnes (1986, p. 7) señalan, apenas si se ha utilizado la aproximación etnográfica en la investigación didáctica, de forma que la retroacción a la actuación ha procedido en su gran mayoría a través de las evaluaciones dadas por los estudiantes utilizando sistemas o escalas de valoración.

Antes de adentrarnos en consideraciones en torno a la evaluación de la actuación conviene apuntar, siquiera someramente, algunas notas sobre las aproximaciones señaladas en la evaluación y las que propugnamos. Sin duda alguna, el modelo evaluativo basado en objetivos ha sido el prevalente. Sin ahondar ahora en la persona que selecciona los objetivos, el propósito de esta aproximación ha consistido en determinar si los objetivos se han conseguido y así esclarecer si ha habido éxito o no en el esfuerzo. Frente a las aproximaciones orientadas a la decisión o al consumidor parecen ser evaluaciones verdaderas, en cuanto que ayudan a los profesores a juzgar y/o mejorar el valor y/o mérito de su enseñanza. Esta última aproximación, debida a Scriven, ha favorecido la aplicación de los sistemas evaluativos con un claro propósito formativo (Stufflebeam y Skinkful, 1985).

La evaluación de la enseñanza: algunas ideas fundamentales

Una de las primeras consideraciones que se debe tener presente es la definición del objeto que se va a evaluar. Si queremos medir la claridad de la enseñanza universitaria debemos precisar qué serie de conductas constituye este constructo. Por tanto, uno de los problemas que afrontan quienes delimitan sistemas es determinar la naturaleza y posterior forma (de conocimiento, de actuación, etc.) que han de adoptar las competencias que incluyan en un sistema. Dado que los inventarios son sistemas finitos de conductas, debe también considerarse cuáles son las fuentes de derivación de los ítems. Como ya señalábamos, los estudios descriptivos, correlacionales y experimentales sobre la enseñanza han sido fuentes para el diseño de sistemas de evaluación sumativa y formativa para profesores (Borich y Fenton, 1977).

Cuando aludimos a estas dos modalidades evaluativas, omitimos la diagnóstica como la tercera de las categorías de aplicación de los sistemas. Bajo esta última, nos preocupa evaluar a los profesores para tomar decisiones sobre su posible contratación, para señalar qué tipo de formación inicial les corresponde. No cabe duda que podría ser una de las modalidades a utilizar con aquellos profesores universitarios que acceden por primera vez a la docencia como profesores asociados. A la luz de los resultados de las evaluaciones se podrían diseñar talleres de perfeccionamiento que facilitarían la consecución de las conductas previstas y no alcanzadas en los sistemas evaluativos. De la misma forma, los análisis formativos proporcionan información específica que facilita la retroacción para la identificación y posterior corrección de conductas débiles en los profesores. La evaluación sumativa acumula los juicios sobre el profesor incorporando diversas variables de actuación. Frente a la modalidad evaluativa anterior, que se apoya en piezas de información discretas y frecuentes, la evaluación sumativa se realiza infrecuentemente dado que con ella se trata de establecer el mérito docente o certificar las destrezas y habilidades que tienen los profesores.

Los estatutos de las distintas universidades que han incluido artículos sobre el control de la calidad de la enseñanza no han clarificado el sentido que tiene la evaluación, permitiendo con esa indefinición generar ansiedad y confusión en los participantes. En este artículo, mostramos nuestra posición que la evaluación formativa debe ser una fase de un proceso de formación del profesorado universitario. La enseñanza superior adolece, en general, de falta de mecanismos de apoyo para su propio perfeccionamiento. Las disposiciones jurídicas creadas para hacer un efectivo control social de la enseñanza no deben enturbiar un derecho que se le debe prestar al profesorado, cual es recibir retroacción de su enseñanza que le afiance como un profesional de la docencia, y como tal, que su trabajo esté abierto a recibir evaluación de estudiantes por medio de instrumentos en los que él ha participado para su elaboración y cuyos fundamentos él mismo comparte. Los atributos o características de un sistema de evaluación eficaz, como nos indica Conley (1987), se deben aplicar a cualquier modalidad evaluativa que se aplique en nuestras universidades.

Indicadores de calidad para evaluar al profesorado universitario

La evaluación del profesor universitario en nuestro país se realiza en dos tiempos o, cuando más, en tres situaciones: una para conseguir la titularidad, que da seguridad en el empleo y otra para obtener la cátedra. Este esquema se incrementa con otra fase para quienes acceden como primer empleo a la categoría de profesor contratado. En otros países, como en Estados Unidos, consta de cuatro tiempos a lo largo de la carrera docente, variando, no obstante, según la naturaleza de los departamentos. En nuestro contexto, no se han estudiado los efectos de los nuevos sistemas de evaluación del profesorado, a pesar de hacerse públicos los criterios para juzgar a los candidatos.

Los indicadores de calidad que se analizan para la selección del profesorado (investigación y docencia) no se corresponden con los sistemas de evaluación que previamente hemos citado, con lo que nos hallamos ante una dicotomía procedimental a la hora de determinar bajo qué principios se debe juzgar a un profesor. Si la concesión de estabilidad administrativa fuese el efecto de una modalidad de evaluación, ésta tendría el carácter de sumativo o cuando menos diagnóstico, y las fuentes de datos que se recopilasen en torno al docente habrían de ser tales que garantizaran que el profesor se ajusta a los valores y necesidades previstas por cada institución universitaria. No es así, y por ello nos encontramos con indicadores distintos a los que se van a utilizar posteriormente para controlar el rendimiento en la enseñanza universitaria.

Donald (1984) ha hecho un resumen de posibles indicadores de calidad a nivel universitario que bien puede constituir una base o fuente para tomar decisiones a la hora de evaluar al profesor bien sea para controlar su rendimiento bien para establecer un standard de calidad que los docentes deben reunir a fin de promoverse en su carrera profesional. Tres son las áreas que incluye: investigación, docencia y servicios. El análisis de cada una de las áreas nos induce a reflexionar sobre algunos criterios a los que no se presta suficiente atención. Así, por ejemplo, bajo el epígrafe de investigación, las veces que un investigador ha sido citado, la calidad de la investigación juzgada por compañeros o las ayudas institucionales concedidas a la investigación. Bajo el área de enseñanza, Donald reconoce la complejidad del fenómeno «buena enseñanza» y de su medida. Llama la atención sobre los métodos usados para evaluar, sobre todo por parte de colegas y aporta hallazgos de investigación que demuestran que la productividad investigadora no se relaciona con la eficacia docente para indicar que los contratos de mejora en la enseñanza en combinación con el autoanálisis son estrategias eficaces. Más compleja es la evaluación de los servicios prestados a la universidad, a su propia profesión y a la comunidad. ¿Cómo se usan los indicadores de calidad para medir la eficacia del profesor? Existen distintas propuestas que combinan criterios derivados de las tres áreas mencionadas otorgando

pesos diferentes a los criterios seleccionados. Enumerarlos aquí sería prolijo. Sin embargo sí es pertinente apuntar una reflexión final. El profesor necesita encontrar en la evaluación un proceso por el cual se sienta satisfecho, perciba una mejora docente y le gratifique

Evaluación de la actuación docente: dos aproximaciones

Como otra muestra de la falta de consenso sobre el concepto de excelencia en enseñanza universitaria citamos la obra de Braskamp y Brandebury (1984) quienes señalan que para definir la buena enseñanza habría que reflexionar en torno a una serie de factores que se incluyen en las siguientes áreas: input, proceso y producto. Como ejemplo los factores del proceso estarían compuestos por los siguientes elementos: ambiente de clase, conductas docentes, actividades de aprendizaje del estudiante, organización del curso y procedimientos de evaluación. Algunas de estas dimensiones procesuales – ambiente de clase y conductas docentes – merecerán un tratamiento en profundidad en el artículo. Ahora sólo queremos resaltar el hecho de que la evaluación de la buena enseñanza en base a los factores de cada área podrían ser el objetivo de un proyecto de investigación, pero difícilmente servirían para el propósito de mejorar la enseñanza individual de cada profesor, para que los alumnos pudieran seleccionar cursos optativos en base a los resultados de las evaluaciones o siquiera para adoptar decisiones profesionales sobre adquisición o promoción en un status administrativo. De ello se desprende que la evaluación de la actuación no es una tarea sencilla y que se debe clarificar cómo un programa evaluativo debe conducir a situaciones de recompensa.

Dada la complejidad que entraña definir la naturaleza de la evaluación, se debe recomendar procedimientos alternativos de evaluación que capten diferentes dimensiones de la enseñanza y que garanticen la calidad técnica de la evaluación. La interpretación que se haga de los resultados se debe unir a estrategias de autodesarrollo docente, a fin de que el profesor evaluado perciba cierta utilidad en el proceso.

Este sentimiento de que se debe prestar más atención al desarrollo del profesor como consecuencia de la evaluación de la actuación es la razón que dio pie a que Stiggitt y Bridgeford (1985) hayan buscado estrategias alternativas a las prácticas evaluativas referidas a profesores no universitarios dado que las existentes eran consideradas frustrantes desde distintas instancias. De ahí el esfuerzo por construir un ambiente más favorable de la evaluación. En su artículo citan el modelo evaluativo de Manatt (Stiggitt y Bridgeford, 1985, p. 87) que puede ayudarnos a reflexionar sobre el diseño de un modelo para la enseñanza universitaria. Como consecuencia, trasladaremos sus recomendaciones a la enseñanza superior: a) se debe implicar a los profesores universitarios en el proceso de evaluación; b) se deben desarrollar criterios de actuación basados en la investigación en las prioridades de cada universidad; c) se deben marcar unas metas en la evaluación; d) hay que procurar la concurrencia de métodos multidimensionales para evaluar destrezas de los profesores, que incluyan desde la recopilación de datos objetivos – autoevaluación y evaluación de colegas; e) analizar los resultados con cada profesor para establecer de común acuerdo metas de perfeccionamiento, y f) inclusión de conferencias de preobservación para la adquisición de datos generales que sirvan de antecedentes de actuación. Como se observa, estamos proponiendo un modelo de evaluación que recuerda la supervisión clínica, que ya hemos ensayado en esta universidad y que proponemos más adelante (Villar, 1983).

El vídeo: ¿Sustituto de una observación en vivo? Cuatro perspectivas en el análisis de una lección

Hay un acuerdo unánime que indica que se deben utilizar múltiples métodos para conseguir una información fiable sobre la actuación del profesor. Entre esos métodos

citan la observación y la evaluación por los estudiantes. En este epígrafe revisaremos el primero.

La observación de la conducta del profesor tampoco está exenta de controversias, que citaremos con ocasión de un simposio celebrado en el Congreso de la American Educational Research Association (AERA) en San Francisco en 1986. La primera de ellas es *qué observar*. Si tenemos que confiar en evaluadores éstos se tendrán que entrenar adecuadamente a fin de que las descripciones conductuales sean válidas y correctamente interpretadas. Una muestra de la conducta del profesor tras una observación es insuficiente para que el profesor y el evaluador –aplicando un criterio previamente acordado– generen información diagnóstica relevante. La información debe ampliarse y utilizar recursos tecnológicos –como el vídeo– para multiplicar las piezas de evidencia. Recopiladas series de actuaciones en un vídeo se debe proceder ahora a su análisis. Es aquí cuando se multiplican los problemas, porque no hay un consenso sobre cuál puede ser la mejor estrategia al respecto. Veamos qué ocurrió en el Simposio de la AERA, (ver *Journal of Teaching and Teacher Education*, 1986).

El Ministro de Educación de Estados Unidos impartió una clase de Estudios Sociales a alumnos de Enseñanzas Medias que fue grabada en vídeo y diseccionada por cuatro expertos en análisis de interacción de clase. Los cuatro analistas seleccionaron cuatro enfoques diferentes: así mientras que B. Rosenshine la analizó en términos de los hallazgos recientes sobre la enseñanza eficaz, Peterson y otros lo hicieron utilizando un sistema que contiene seis instrumentos observacionales (cinco formativos y uno sumativo) y que se ha usado ampliamente en Florida después de haberlo sometido a pluralidad de análisis psicométricos. La tercera vía –un análisis sociolingüístico del discurso– fue aplicada por Shuy en base a cinco rutinas: marco, secuencia interrogativa, tópico, respuesta, y análisis de evaluación. Finalmente, la visión de Eisner utilizando sus genuinos métodos del «connaissance» educativo y de la crítica educativa mostró una imagen del profesor-ministro que bien parecía una viñeta narrativa. Aún más. La investigadora S. Delamont desde una tradición británica criticó los cuatro análisis para afirmarse en una posición etnográfica. El profesor-ministro reaccionó más favorablemente ante la descripción de E. Eisner, no tanto por su análisis científico cuanto por su valoración humanística.

La observación realizada por los colegas suele estar sesgada. Si no adiestramos a los observadores en aproximaciones de observación y de análisis, no resultará extraño que después no acierten a comunicar los resultados a los profesores. El entrenamiento de los observadores se deberá hacer en aquella estrategia que mejor responda a las metas previstas por la institución para el proceso evaluativo. Una constante de este artículo es que la universidad debe proveer cursos para el desarrollo del profesorado y para el entrenamiento de los observadores.

Evaluación de la enseñanza universitaria por los estudiantes

La evaluación de la enseñanza por los estudiantes es el método más difundido para valorar la actuación de los profesores universitarios y también la recomendación política que figura en los textos estatutarios de las distintas universidades.

De nuevo, intentaremos seleccionar algunas controversias que presiden la utilización de esta estrategia a fin de recomendar a futuros evaluadores algunos principios que deben tener presente. No insistimos en los propósitos de esta modalidad evaluativa, porque coinciden con los generales establecidos por la institución. De modo que incidiremos en los problemas metodológicos inherentes al diseño y construcción de instrumentos, y en los tipos de investigación basados en ellos.

Marsh (1984) a la hora de diseñar y desarrollar su «Students' Evaluations of Educational Quality» (SEEQ) siguió una validación de construcción, que bien puede ser una referen-

cia para futuros investigadores que están iniciando el desarrollo de cuestionarios. Alg unos supuestos que originan la investigación sobre la evaluación por los estudiantes reflejan preocupación con que hemos iniciado este artículo: que la eficacia docente multidimensional y como consecuencia los instrumentos deben comprender declaraciones que la expliciten. Las declaraciones o ítems del cuestionario se tendrán que expresar en tales términos que representen la enseñanza que se va a evaluar. En este sentido, las dimensiones del cuestionario se deben promover haciendo un análisis lógico de los grupos de ítems y posteriormente comprobarlas empíricamente mediante un análisis factorial. Marsh (1984, p. 710) destaca la importancia de la retroacción diagnóstica cuando un instrumento se compone de dimensiones separadas. Los instrumentos deben responder con coeficientes altos a la medición de la fiabilidad. Si bien el problema más complejo a medir es la generalidad de las puntuaciones a lo largo de cursos y profesores, un problema que parece insoluble es el de la validez de los instrumentos que miden la buena enseñanza dada la ambigüedad del término, y complejo parece ser el diseño de una investigación multiseccional que valide un instrumento utilizando el aprendizaje del estudiante medido a través de exámenes objetivos. Finalmente, llamamos la atención sobre la hipótesis del sesgo de las características generales de los estudiantes que afectan sus evaluaciones, como por ejemplo, interés anterior por la asignatura, notas esperadas y Tarea/Dificultad.

La evaluación por los estudiantes ha sufrido críticas diversas que ha rebatido Aleamoni (1987). Una de ellas es que las evaluaciones por los estudiantes no son fiables. Sin embargo, se ha demostrado que sí lo son, cuando, por el contrario, las evaluaciones dadas por los colegas carecen de ello. Otro de los mitos es que los estudiantes no son capaces de juzgar adecuadamente a sus profesores hasta que no han abandonado la universidad ahí, que algunos estatutos, preconicen que evaluarán la calidad de la enseñanza a los alumnos egresados de la universidad. Esta recomendación no parece apoyarse en evidencias empíricas, y sí más bien en el supuesto racional de que sus evaluaciones serán válidas y fiables, por cuanto están menos expuestas a los efectos del sesgo de las características generales, y quizás también, porque se acepte el mito 14 de Aleamoni (p. 114) que señala que «las notas o calificaciones que reciben los estudiantes en el curso se correlacionan altamente con las evaluaciones del curso y del profesor».

La aplicabilidad de instrumentos en distintos contextos

No resulta extraño que un considerable número de cuestionarios sean americanos cuando se analiza la dilatada tradición de ese país en esa práctica evaluativa. Como consecuencia de una de las vías de aproximación a la evaluación de la enseñanza universitaria por los estudiantes, sería mediante la utilización de instrumentos previamente fiabilizados, y no de constatar su validez en otro marco institucional. En un momento histórico en el que en nuestro país se está iniciando en esa corriente de estudio, los investigadores podrían desarrollar un vademecum con aquellos instrumentos de probada fiabilidad para aplicarlos en nuestro país ofreciendo similares resultados, evitándose el farragoso proceso de desarrollo de un cuestionario, cuando las metas institucionales y el constructo a medir fuesen análogos.

Marsh, Touron y Wheeler (1985) propusieron un estudio tendiente a comprobar la aplicabilidad en España de dos instrumentos americanos que se habían utilizado anteriormente en Australia (el Instrumento Endeavor y el Instrumento SEEQ). La muestra de estudiantes procedía de la Universidad de Navarra. Uno de los propósitos más importantes que tenía el estudio, en palabras de los autores, era determinar si los componentes de la eficacia docente identificados por los estudiantes americanos se podían corroborar a través de las respuestas de los estudiantes españoles. Los resultados parecen confirmar que los instrumentos americanos se pueden aplicar al marco español y que los componentes que estructuran el concepto de eficacia docente en las universidades americanas son los mismos que en el caso español.

Estos dos instrumentos fueron aplicados posteriormente en la Universidad de Canterbury en Nueva Zelanda (Watkins, Marsh y Young, 1987), a fin de comprobar la misma hipótesis de trabajo desarrollada en el caso de la Universidad Navarra. Ensayados los instrumentos en cuatro países, bien puede hablarse de la validez intercultural de estos instrumentos para medir un modelo de eficacia docente.

Otra perspectiva es la que adoptaron Aparicio, Tejedor y Sanmartín (1981) a fin de diseñar un cuestionario para la Universidad Autónoma de Madrid, en una fecha anterior a esta reciente preocupación de evaluar al profesorado universitario. Como fuente para la elicitación de sus primeros 132 ítems seleccionaron cuatro cuestionarios elaborados en los años setenta en Estados Unidos. El cuestionario final tenía 60 ítems. La solución factorial practicada sobre los ítems permitió seleccionar como dimensiones de eficacia docente los factores: Entusiasmo/Dominio de la asignatura; Organización/Claridad/Preocupación por la enseñanza; Interés por el alumno individual; Interacción con el grupo/Preocupación por los resultados, y Dedicación.

Obsérvese cómo al aplicar los instrumentos Endeavor y SEEQ en Estados Unidos, Australia y España, los mayores pesos fueron en las dimensiones Entusiasmo docente, Aprendizaje/Valor y Claridad de presentación.

Es decir, que a través de estos instrumentos aparecen dimensiones de eficacia docente que luego hemos corroborado en nuestro propio estudio del DIES, que describimos en el epígrafe siguiente.

Estudio de caso: El DIES

La investigación «Diagnóstico Instruccional en la Enseñanza Superior» (Villar, 1983a) es un ejemplo de sistematización del desarrollo del profesorado universitario. En ese programa se integran distintas fuentes para el diagnóstico de la actuación del profesor que ya hemos citado anteriormente, como son el uso del vídeo para el registro de las actuaciones en clase y la evaluación por los estudiantes. El programa combina, pues, la evaluación con el desarrollo del profesorado, y la mejora se hace siguiendo el formato de la supervisión clínica. Como indicamos anteriormente, al citar las recomendaciones de Manatt, el DIES incorpora criterios de evaluación basados en la investigación, por medio del cuestionario Análisis Docente por los Estudiantes (ADE), se explicitan las metas del programa de perfeccionamiento, se incluyen métodos multidimensionales para la evaluación de la actuación, se establecen metas de perfeccionamiento basadas en las decisiones de los profesores, y se prevén varias conferencias de preobservación y de retroacción a lo largo del proceso.

Quisiera, no obstante, abundar más en el cuestionario ADE, dado que estamos analizando ahora esa fuente como modalidad para la evaluación de la conducta docente. El ADE fue una adaptación de un cuestionario americano que sintetizaba la corriente de investigación dominante en los años setenta sobre eficacia docente y evaluación de la enseñanza universitaria por los estudiantes. Incluía ítems, pues, que procedían de otros instrumentos y destrezas que se habían ensayado en programas de microenseñanza. Así, en el ADE se declaran algunas destrezas y habilidades similares a las de los instrumentos que hemos citado en el epígrafe anterior, como el entusiasmo o la claridad del profesor. Con el ADE llegamos a comprobar que los profesores de la muestra española tenían unas competencias didácticas similares a las de otra muestra de profesores americanos evaluados con el mismo instrumento. El cuestionario fue la base para la derivación de objetivos de entrenamiento de un programa para remediar las debilidades docentes percibidas por los profesores (Villar, 1983b). En este sentido, otros instrumentos diseñados de la misma forma han sido orientados a la evaluación formativa de los profesores.

Antes citamos el «Florida Performance Measurement System». Ahora revisaremos el «Connecticut Teaching Competencies» (Streifer e Iwanicki, 1987) para señalar la pléora

de instrumentos contruidos siguiendo la filosofía de evaluar la actuación a partir de la tradición de la eficacia docente. El uso que se dio a este instrumento fue garantizar la calidad de los docentes que entraban en la profesión. Por consiguiente el sistema guía el proceso que comienza en la formación inicial y finaliza con la certificación de los profesores. Llamamos la atención sobre la validación que se hace de este instrumento que admite decisiones metodológicas emanadas del mismo Ministerio de Salud, Educación y Bienestar, consistentes en la utilización de la técnica Delfos para diseñar y refinar inicialmente el cuestionario ante un número limitado de educadores. En la segunda fase se consultó a través de un muestreo aleatorio estratificado y proporcional a profesores de los niveles universitarios y a directores de centros educativos. Las técnicas usadas en el diseño de esta medida de la actuación docente pueden ser útiles cuando se quiera elaborar un instrumento similar -en términos de la misma aproximación didáctica- para el nivel universitario.

Evaluación del ambiente de aprendizaje

En la enseñanza superior han florecido instrumentos que han intentado medir el clima de clase, departamento o institución universitaria. Como en los instrumentos que miden la eficacia docente, los estudiantes han sido los agentes de este tipo de evaluación. El impacto del ambiente o clima de clase en el proceso didáctico, a pesar de ser abundante la literatura que ha hipotetizado su importancia, no se ha constatado suficientemente. La metodología seguida para identificar los elementos constituyentes del ambiente también ha variado respecto a la aproximación de la eficacia docente. Ahora se presta más atención a las técnicas etnográficas (que se desarrollaron a comienzos de los años setenta) para identificar las impresiones que perciben los alumnos en torno a este constructo.

Winteler (1981) ha seleccionado el departamento como unidad ambiental que facilita la enseñanza-aprendizaje, y ha revisado estudios hechos en Holanda, Inglaterra y Alemania para informar de las características ambientales que deben poseer los departamentos desde las perspectivas de alumnos y profesores. Su revisión no es sólo descriptiva sino que muestra ejemplos de estudios que asocian las características ambientales con los criterios más divulgados a la hora de determinar la eficacia del ambiente: esto es satisfacción y el rendimiento del estudiante.

Si el clima se puede detectar y describir; si el clima se asocia con criterios deseables y se conoce cómo cambiarlo ¿no nos hallamos ante una línea de investigación en la que puede existir mayor consenso y aceptación entre los educadores, y que puede repercutir más satisfactoriamente en el profesorado? En una reciente revisión de la literatura sobre la contribución de la teoría necesidad-presión, Genn (1984) ha sistematizado dicha literatura -prioritariamente australiana- llegando a declarar que este «procedimiento dado que no amenaza y es aparentemente valioso, puede contribuir significativamente al autoconocimiento y comprensión de una institución». Un ejemplo lo tenemos en el estudio de un componente formativo tan difuso como son las prácticas de enseñanza de las EE.UU. Formación del Profesorado que describimos seguidamente.

La evaluación del ambiente es una línea de investigación que también afecta a los niveles no universitarios, en donde existe una pluralidad de sistemas para medir el clima de clase. Citamos ahora esta corriente porque hay ejemplos de investigación que incluyen a los alumnos de los niveles no universitarios para evaluar profesores en formación. El estudio de Duschl, Waxman y Morecock (1986) puede servirnos de referencia. Aproximadamente 500 alumnos de enseñanzas medias completaron el instrumento «Individualized Classroom Environment Questionnaire» (ICEQ), cuyas marcas se combinaron con las aportadas por otros grupos de codificadores (supervisores universitarios, profesores tutores de centro y profesores en formación). Los autores informan que las evaluaciones de los alumnos pueden ser útiles para asistir en el proceso de evaluación de las prácticas de enseñanza.

Estudio de caso: El «inventario de ambiente de clases universitarias» (IACU)

Los escasos ejemplos de instrumentos para medir el clima o nivel universitario ni han ahondado suficientemente el marco de clase como foco de estudio ni se han centrado en las EE.UU. de Formación del Profesorado como instituciones específicas de nivel superior.

Nuestro estudio se propuso diseñar un instrumento para medir el ambiente de clase en EE.UU. de Formación del Profesorado de E.G.B. del distrito universitario de Sevilla, describir el ambiente de clase en este tipo de centros y comparar los ambientes de clase en base a ciertas variables seleccionadas.

El procedimiento seguido para diseñar el cuestionario se apoyó en dos estrategias etnográficas como son la entrevista y la observación participante. Seleccionada una E.U. de Formación del Profesorado como si se tratara de un estudio de caso, se entrevistaron a profesores y a estudiantes de dicho centro y se observaron las clases impartidas por los profesores. Las entrevistas fueron grabadas en audio y posteriormente transcritas, a fin de derivar declaraciones sobre el ambiente o clima de clase. Los observadores prestaron atención a aquellos hipotéticos elementos que calificaran el clima de las clases observadas.

El banco de declaraciones fue tan exhaustivo que se utilizaron técnicas estadísticas multivariantes para reducir el número de elementos del instrumento a un tamaño tal que fuese fácil su administración y corrección, y que describiera suficientemente el clima de estos centros. En el proceso de reducción del cuestionario se implicaron a profesores de EE.UU. de Formación del Profesorado del territorio español, y a estudiantes de la E.U. de Magisterio de Sevilla, que dieron lugar después de dos versiones provisionales a un cuestionario de 49 declaraciones y siete dimensiones. Además, se propusieron dos versiones del IACU manteniendo las mismas subescalas. Estas dimensiones (Cohesión, Satisfacción, Personalización, Orientación a la tarea, Innovación, Evaluación y Gestión de clase) son los indicadores del ambiente y en base a ellos se compara el clima de clase. En el cuadro que figura a continuación mostramos una descripción del instrumento que sigue la tradición originada en Moos, y que se apoya en su conceptualización y medida en la teoría necesidad-presión (Fraser, 1986).

El instrumento precisa investigación adicional. Sería conveniente aplicarlo en otras EE.UU. del Profesorado para conocer si se mantiene la estructura diseñada de ítems y subescalas, conocer si el clima provoca una diferencia en resultados esperados o si se puede cambiar aquellas dimensiones que se conozca se asocian negativamente con un criterio deseado. Finalmente, habría que buscar antecedentes y causas que establecen las diferencias en las distintas escalas cuando se comparen aulas dentro de una misma escuela o entre escuelas.

No obstante, la evaluación del clima de clase puede ser una vía alternativa a la evaluación de la eficacia docente. Desde un punto de vista experimental, los profesores han aceptado con mejor grado la evaluación por los estudiantes de un constructo intangible como es el clima de su conducta docente (Villar, 1986 y 1987).

Uso de la evaluación docente y de las percepciones ambientales para cambiar contextos de clase

Una de las preguntas clásicas en la investigación sobre el clima de clase se debió a Stern quien la formuló en los siguientes términos: «¿Qué constituye un ambiente óptimo, uno que satisface o uno que estimula?» (citado en Genn, 1984). Los cuestionarios diseñados recientemente, como ya hemos indicado con el IACU, incluyen dos versiones que miden el ambiente real y el ideal o preferido. Las dos versiones han facilitado la propuesta de hipótesis, que interaccionan variables referidas a las características personales con los

CUADRO
INFORMACION DESCRIPTIVA DEL «INVENTARIO DE AMBIENTE DE
CLASES UNIVERSITARIAS»

ESCALA	CATEGORIA DE MOOS	DESCRIPCION	EJEMPLO Y SIGNIFICADO
COHESION	Relación	Grado en que los estudiantes se conocen y son amigos entre sí.	Entre la mayoría de los miembros de esta clase existe una relación de confianza (+).
SATISFACCION	Relación	Grado en que existe diversión con el trabajo de clase.	En esta clase, alumnos no escuchadas y ideas (+).
PERSONALIZACION	Relación	Grado en que se dan oportunidades al alumno para interactuar con el profesor, y preocupación por el bienestar personal del alumno.	El profesor con el buen juicio de los alumnos (+).
ORIENTACION A LA TAREA	Desarrollo personal	Grado en que es importante completar y permanecer en la asignatura.	En esta clase, se simulan situaciones escolares y se resuelven sus problemas (-).
INNOVACION	Cambio y mantenimiento en el sistema	Grado en que un profesor planifica actividades y técnicas nuevas, infrecuentes y variadas, y anima a los estudiantes a que programen, y a que piensen creativamente.	En esta clase, alumnos inventan y componen con las tareas sugeridas por el profesor (+).
EVALUACION	Cambio y mantenimiento en el sistema	Grado en que un profesor explicita las normas de control instrumental de los estudiantes.	En esta clase, alumnos conocen perfectamente los objetivos de la asignatura (-).
GESTION DE CLASE	Cambio y mantenimiento en el sistema	Grado en que un profesor ordena, organiza y da normas sobre la instrucción de clase.	En esta clase, profesor informa siempre a los alumnos de los procedimientos que servirán para desarrollar la enseñanza (+).

tipos de ambiente medidos por las dos versiones de los cuestionarios, y, en particular, si los resultados afectivos y cognitivos son mejores cuando hay un ajuste entre ambiente real y preferido (Fraser, 1986). No obstante, son escasos los estudios realizados que muestren cómo cambia el ambiente y qué dirección adopta el cambio.

En la investigación sobre la conducta del profesor, Stodolsky (1984) nos llama la atención sobre los problemas existentes a la hora de generalizar la conducta docente. Los estudios sobre la generalización se cuestiona si las medidas o estimadores de la conducta son estables a lo largo de varias condiciones de medida. En general, las condiciones que se varían son los evaluadores, los sistemas (de alta o baja inferencia), asignatura que se explica y ocasiones de las observaciones. Los datos que se conocen hasta ahora no permiten concluir que se conozca suficientemente qué estabilidad tiene la conducta docente a lo largo de las distintas situaciones, y como éstas pueden provocar diferentes estimadores conductuales. ¿Qué decisiones sobre la conducta del profesor se pueden adoptar a la luz de los resultados de evaluación? Scriven (1987) critica la aproximación a la evaluación docente conocida por «evaluación del profesor basada en la investigación» porque ni se define el criterio, ni se indica el grado en que las conductas seleccionadas son generalizables a otros contextos.

Teniendo presente estas limitaciones conceptuales y metodológicas, ¿cómo se puede optimizar el proceso de evaluación que conduzca a los profesores a modificar su conducta de clase o algunas dimensiones ambientales que se correlacionan significativamente con resultados de aprendizaje? Una propuesta que vamos a sugerir ahora se debe a Fraser (1986) y que luego hemos reelaborado (Villar, 1987b). Los perfiles ambientales de clase combinando los climas real y preferido permiten hallar similitudes y congruencias que pueden despertar curiosidad en el sujeto que hace autoevaluación o cuando la evaluación es asistida con la ayuda de un preparador o consultor. El profesor, suponemos que hallará sorpresa en determinados resultados y que la reflexión sobre los mismos serán los fundamentos para mejorar su actuación. La reflexión como base para la mejora de la actuación y el cambio curricular ha constituido una metodología que ha arraigado en contextos anglosajones, y que se está iniciando también en nuestro país.

Desarrollo del profesorado

La evaluación de la actuación y del clima de clase se deberían utilizar como técnicas de retroacción alternativas que identifiquen áreas en las que un profesor difiere de su ideal. El uso de instrumentos como recursos de retroacción se ha utilizado en programas de formación del profesorado basados en la actuación, como el análisis de interacción, la microenseñanza o la supervisión clínica. En la revisión que hacen Levinson-Rose y Menges (citados en Dunkin y Barnes, 1986, p. 772) de las aproximaciones a la mejora en la enseñanza superior están las siguientes: ayudas para el desarrollo de proyectos, talleres y seminarios; microenseñanzas y minicursos, y entrenamiento en conceptos usando protocolos de incidentes docentes. Como se observa, el desarrollo del profesorado ha combinado estrategias basadas en una psicología comportamental, en propuestas de programas convencionales y en subvenciones administrativas.

Desde nuestra experiencia como formador de profesores, se tiene que extremar el diseño y la aplicación de programas de formación del profesorado universitario que tengan significado para los docentes. Los profesores deben percibir la «interacción» que existe entre el programa y el cambio personal y socioorganizativo de la universidad en donde se inserte. Debe percibir que el programa provee estímulos que premian e incentivan su carrera docente. Si definir la naturaleza del programa es una labor compleja, no lo es menos explicitar todas las cuestiones que le dan entidad, como son por qué, dónde, cuándo, cómo y qué constituye el programa formativo. A esta nota de «comprensividad» se añade la de «continuidad». Es decir, que la unidad de desarrollo del profesorado no

establezca un cursillo como vehículo para el cambio de la conducta docente o del clima de la clase, sino que se faciliten actividades individuales de seguimiento que frecuenten oportunidades de retroacción. La «potencia» del programa tiene que ver con el significado que le atribuyan los participantes. Si en él se observa la practicalidad del entrenamiento el programa tendría relevancia. Por ello la combinación de instrumentos de evaluación con otras técnicas etnográficas –como la entrevista con el preparador– puede multiplicar las evidencias para mejorar una actuación. Cuando citamos a un preparador o a una grabación en vídeo de una clase, es porque la unidad de desarrollo formativa debe proporcionar «soporte material y personal» cualificado para el perfeccionamiento del profesor. Finalmente, se tiene que «documentar» todas las fases del proceso formativo incluida la constatación de los cambios habidos como resultado del programa (Howe y Vaughan, 1983). No es fácil elaborar un programa que responda a estas características si quiera que se inspire en modelos de investigación y desarrollo, como el ensayo de Friedmann (1987), aunque su primera fase para la propuesta de metas tuviese como origen el controvertido paradigma proceso-producto. Sin embargo, si resaltamos como adecuada su tercera y última fase, que es el análisis de vídeos con la ayuda de un supervisor.

El desarrollo del profesorado universitario debe ser individualizado. Por ello, sugerimos el establecimiento de relaciones diadas entre un profesor y un supervisor o asesor. «coaching» (preparación) técnico puede ser una estrategia formativa de gran utilidad en la enseñanza superior. Este proceso multiplica las oportunidades de retroacción (nótese nuestra insistencia en incrementar la frecuencia de ocasiones de retroacción) en ambientes donde existen relaciones de ayuda (fundamento humanista de entrenamiento). Después «coach» o preparador tendrá que indicar cómo se presenta un contenido con demostración (esto implica que existan formadores dentro del departamento o que la formación sea en pares de colegas), que se hagan muchas prácticas y que el entrenamiento tenga lugar en el marco de trabajo. El aula puede servir, entonces, como fuente de información y como soporte para el desarrollo del profesorado (Brandt, 1987a).

BIBLIOGRAFIA

- ALEAMONI, L.A. (1987). Student Rating Myths Versus Research Facts. *Journal of Personnel Evaluation in Education*, 1 (1), 11-119.
- APARICIO, J.J., TEJEDOR, F.J. y SANMARTIN, R. (1981). *La enseñanza universitaria vista por los alumnos: Un estudio para la evaluación de los cursos en la enseñanza superior*. Madrid: ICE de la Universidad Autónoma.
- BORICH, G.D. and FENTON, K.S. (1977). *The Appraisal of Teaching: Concepts and Process*. Reading: Addison Wesley Publishing Company.
- BRANDT, R.S. (1987a). On Teachers Coaching Teachers: A Conversation with Bruce Joyce. *Educational Leadership*, 44 (5), 12-17.
- BRANDT, R.S. (1987b). On Teacher Evaluation: A Conversation with Tom McGre. *Educational Leadership*, 44 (7), 20-24.
- BRASKAMP, L.A., BRANDENBURY, D.C. and ORY, J.C. (1984). *Evaluating Teaching Effectiveness. A Practical Guide*. Beverly Hills: Sage Publications, INC.
- CONLEY, D.T. (1987). Critical Attributes of Effective Evaluation Systems. *Educational Leadership*, 44 (7), 60-64.
- DELAMONT, S. (1986). Discusión: A View from a Quadruple Outsider, *Teaching and Teacher Education*, 2 (4), 329-332.
- DONALD, J.G. (1984). Quality indices for faculty evaluation. *Assessment and Evaluation in Higher Education*, 9 (1), 41-52.

- DUNKIN, M.J. and BARNES, J. (1986). Research on Teaching in Higher Education. En WITTRICK, M.C. (Ed.). *Handbook of Research on Teaching, Third Edition*. N.York: Macmillan. 754-777.
- DUSCHL, R.A., WAXMAN, H. and MORECOCK, R. (1986). *A comparison of students', student teachers', cooperating teachers', and university supervisors' perceptions of the science classroom environment*. Paper presented at the Annual Meeting of the AERA, San Francisco.
- ELTON, L. (1987). *Teaching in Higher Education: Appraisal and Training*. London: Kogan Page Ltd.
- FRASER, B.J. (1986). *Classroom Environment*. London: Croom Helm.
- FRIEDMAN, M. (1987). *Teaching Urban College Instructors to Teach Better*. Paper presented to the Annual Meeting of the AERA. Washington.
- GENN, J.M. (1984). Research into the Climates to Australian Schools, Colleges and Universities: Contributions and Potential of Need-Press Theory. *The Australian Journal of Education*, 28 (3), 227-248.
- HOWEY, K.R. and VAUGHAN, J.C. (1983). Current Patterns of Staff Development. En GRIFFIN, G.A. (Ed.). *Staff Development*. Chicago: The University of Chicago Press, 92-117.
- MARSH, H.W. (1984). Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases, and Utility. *Journal of Educational Psychology*, 76(5), 707-754.
- MARSH, H.W., TOURON, J. and WHEELER, B. (1985). Students' Evaluations of University Instructors: The Applicability of American Instruments in a Spanish Setting. *Teaching and Teacher Education*, 1 (2), 123-138.
- SCRIVEN, M. (1987). Validity in Personal Evaluation. *Journal of Personnel Evaluation in Education*, 1 (1), 9-23.
- STIGGINS, R.J. and BRIDGEFORD, N.J. (1985). Performance Assessment for Teacher Development. *Educational Evaluation and Policy Analysis*, 7 (1), 85-97.
- STODOLSKY, S.S. (1984). Teacher Evaluation: The Limits of Looking. *Educational Researcher*, 13 (9), 11-18.
- STREIFER, P.A. and IWANICKI, E.F. (1987). The Evaluation of Beginning Teacher Competencies in Connecticut. *Journal of Personnel Evaluation in Education*, 1 (1), 33-35).
- STUFFLEBEAM, D.L. and SHINKFIELD, A.J. (1985). *Systematic Evaluation. A Self-Instructional Guide to Theory and Practice*. Boston: Kluwer-Nijhoff Publishing.
- VILLAR ANGULO, L.M. (Dir.) (1983a). *Diagnosis instruccional en la enseñanza superior*. Sevilla: ICE de la Universidad.
- VILLAR ANGULO, L.M. (1983b). Evaluación de la enseñanza universitaria por los estudiantes. *Enseñanza. Anuario interuniversitario de didáctica*. Salamanca: Ediciones Universidad de Salamanca. 263-297.
- VILLAR ANGULO, L.M. (1986). De la evaluación de la conducta docente a la evaluación de contextos en la enseñanza universitaria: el caso de las escuelas universitarias del profesorado de la Universidad de Sevilla. *Cuestiones Pedagógicas*, 3, 127-144.
- VILLAR ANGULO, L.M. (1987). Evaluación del ambiente de aprendizaje en el aula universitaria. *Revista de Innovación e Investigación Educativa*, 2, 53-67.
- WATKINS, D., MARSH, H.W. and YOUNG, D. (1987). Evaluating Tertiary Teaching: A New Zealand Perspective. *Teaching and Teacher Education*, 3(1), 41-53.
- WINTELER, A. (1981). The Academic Department as Environment for Teaching and Learning. *Higher Education*, 10, 25-36.