

## MODELOS DE EVALUACION UNIVERSITARIA

por ARTURO DE LA ORDEN HOZ  
*Universidad Complutense de Madrid*

### *La evaluación del aprendizaje de los alumnos*

La educación formalizada en las instituciones escolares supone un conjunto de procesos para estimular el aprendizaje, que se desarrollan en un ambiente o contexto que ofrece a los alumnos la oportunidad de alcanzar ciertos resultados, cuyo logro, a su vez, es comprobado mediante alguna forma de evaluación.

Con este marco de referencia, podemos afirmar que la evaluación del aprendizaje de los estudiantes constituye, sin duda, uno de los procesos nucleares en el contexto de la enseñanza, y de la educación en su conjunto. Formalmente constituye la fase terminal del ciclo completo de la instrucción. El ciclo comienza con el diseño de la intervención (análisis de contenidos, formulación de objetivos y desarrollo de planes y estrategias docentes-discentes y selección y adaptación de materiales). La segunda fase consiste en la realización o ejecución de la intervención diseñada para posibilitar y facilitar a los estudiantes el dominio de los objetivos previstos. Finalmente, el ciclo se cierra con la determinación de si los alumnos han logrado o no los objetivos y, en caso positivo, en qué forma, grado o nivel.

Pero la evaluación, aunque formalmente conceptualizada como la fase final del ciclo instructivo, está permanentemente actuando en todo el proceso. Está presente en el diseño y en la intervención educativa. Puede, de hecho, afirmarse que toda decisión didáctica, orientadora y organizativa se apoya, implícitamente o explícitamente en la evaluación. En otras palabras, la evaluación es un componente intrínseco del proceso educativo en su conjunto, en la medida en que la educación sea una actividad sistemática intencional.

La evaluación supone, a su vez, un proceso, o conjunto de procesos, para la obtención y análisis de la información relevante en que apoyar el juicio de valor, en que esencialmente consiste, como soporte de una decisión. Así pues, en todo acto evaluativo se halla implícito el proceso de cerciorarse del valor de la realidad educativa, según su grado de adecuación (o inadecuación) a una instancia de referencia o criterio. Por tanto, los problemas básicos de la evaluación educativa se centran fundamentalmente en los siguientes puntos:

- Recogida, síntesis, codificación y expresión de la información relevante para describir la faceta evaluada (aprendizaje de los alumnos, en nuestro caso). En otras palabras, supone la determinación de lo que se ha de evaluar (ámbito de la evaluación educativa) y de los procedimientos y formas de evaluación (instrumentos de medida, observación, análisis, etc.).
- Establecimiento y formulación de la instancia de referencia o criterio de evaluación (auto o hetero-evaluación, auto o hetero-referencia, referencia normativa, referencia criterial, etc.). Es decir, precisar los criterios en que el profesor o evaluador ha de apoyar el juicio de valor respecto a las realizaciones de los estudiantes.
- Determinación y clarificación de las decisiones que han de ser adoptadas como resultado de la evaluación. Concretamente, fijar los usos de la evaluación o las alternativas de acción posibles. ¿Para qué evaluamos en cada caso? ¿Para aprobar o suspender a los alumnos?, ¿para mejorar el proceso didáctico?, ¿para otorgar un título?, ¿para...?

*Influencia de la evaluación en el proceso y  
en el resultado de la enseñanza y el aprendizaje*

Parece evidente que si la evaluación del aprendizaje de los alumnos es un elemento o componente esencial del proceso instructivo total, la concepción prevalente de la enseñanza universitaria y las características de su realización determinarán en gran medida la naturaleza y el alcance del modelo evaluativo; pero, a su vez, el modelo de evaluación influirá decisivamente en el proceso y producto de la educación y, en consecuencia, en su calidad. En otros trabajos [1] formulaba lo que podría ser la tesis general, en los siguientes términos: «*Sea cual fuere la modalidad, sistema o tipo de evaluación vigente en un contexto educacional, su influencia sobre el proceso y el producto de la educación es decisiva y afecta directa o indirectamente a su calidad.*»

En primer lugar, constituye un hecho de experiencia común entre los

educadores que los exámenes, lo que se exige y cómo se exige a los alumnos en los exámenes, definen los objetivos reales del aprendizaje y la enseñanza. Calonghi [2] expresa esta idea afirmando que «los criterios de evaluación, pronto o tarde, consciente o inconscientemente, explícita o implícitamente, se constituyen de hecho en los fines de la educación, de la formación cultural». En efecto, uno de los objetivos prioritarios de los estudiantes, en general, es satisfacer las exigencias de los exámenes; y todo profesor conoce la capacidad de los alumnos para percibir las características, hábitos y peculiaridades evaluadoras de los docentes (temas con mayor probabilidad de salir, tipo de cuestiones, formas de calificar, etc.) y para transformar en las metas reales de su aprendizaje los conocimientos y conductas que les permitan superar con éxito las pruebas.

Por otra parte, los propios profesores, especialmente cuando los alumnos han de someterse a exámenes externos a la propia clase o al centro, planifican y desarrollan la enseñanza en función de los criterios de tales pruebas, formulados o no expresamente. Tales criterios se convierten pronto en los objetivos reales de la acción docente y educadora del profesor. La evidencia en este sentido es abrumadora. A este respecto, no puede dejar de constatarse cómo las actuales pruebas de acceso a la Universidad, por ejemplo, han dado un giro radical a los objetivos del curso de orientación universitaria, hasta el punto de que la orientación sólo aparece en el nombre, lo cual no deja de ser altamente desorientador. Por otra parte, los resultados de la investigación sistemática confirman reiteradamente el influjo de la evaluación en el aprendizaje, como se pone de manifiesto en los trabajos de Miller y Parlett [3], Dearden [4], Marton y Saljo [5], Elton y Lanvillard [6] y Fuchs y otros [7], entre otros.

En resumen, el modelo de comportamiento implícito en los criterios de evaluación, al constituir una expresión concreta de lo que se espera del alumno realmente, puesto que es lo que se le exige de hecho, define de una manera efectiva los objetivos reales y operantes de la educación, independientemente de lo estipulado en los programas y de que existan o no formulaciones expresas de los mismos.

En consecuencia, si consideramos el principio de la determinación teleológica de los medios (con toda la relatividad que se quiera, los fines y objetivos condicionan de alguna manera los medios para alcanzarlos), debemos concluir que la evaluación al prescribir realmente los objetivos de la educación, determina, en gran medida, las características de la enseñanza y el aprendizaje, lo que los alumnos aprenden y cómo lo aprenden, lo que los profesores enseñan y cómo lo enseñan, los contenidos y los métodos, en otras palabras, el producto y el proceso de la educación y, en consecuencia, su calidad. En efecto, si la calidad de la

educación está vinculada a un sistema de coherencias en el proceso total de la educación, la virtualidad de la evaluación, como estímulo o freno de la calidad educativa, dependerá enteramente de las características del esquema evaluador. Si los criterios y modos de evaluación, que determinan los objetivos reales y, en última instancia, el producto de la enseñanza y el aprendizaje son coherentes con los objetivos formalmente establecidos —o sobre los que existe un consenso implícito (eficacia)— y, a través de ellos con los fines generales de la educación y el sistema de valores del que derivan (funcionalidad), la evaluación actuará como el más poderoso factor de promoción de la calidad educativa, al garantizar la congruencia y eficacia del sistema en su conjunto. Si, por el contrario, los criterios y modos de evaluación movilizan los procesos educativos hacia objetivos y resultados no coherentes con las metas formalmente establecidas o implícitamente aceptadas, ni en consecuencia, con los fines de la educación y el sistema axiológico en que se sustentan, la evaluación se constituye en el mayor obstáculo a una educación de calidad. Este sería el caso, por ejemplo, de un sistema de exámenes cuya característica esencial fuera exigir la reproducción de información fáctica, contribuyendo así a fijar como objetivo real de la enseñanza y el aprendizaje la asimilación, reconocimiento y evocación de datos más o menos relevantes. Estos objetivos reales canalizarán casi necesariamente las energías y actividad de alumnos y profesores hacia un aprendizaje memorístico y fomentarán los métodos didácticos verbalistas y repetitivos y el abuso del texto entendido como catecismo. Si las metas y objetivos previstos para la enseñanza no coinciden con la memorización de datos e información verbal —supongamos que fueran el desarrollo de la capacidad crítica frente a los hechos y de la capacidad para usar la información aplicándola a la resolución de problemas— es evidente que tal sistema de exámenes constituye un formidable obstáculo a la calidad de la educación. Si por el contrario, el sistema de exámenes se caracterizara por exigir de los alumnos análisis de situaciones supuestas y reales, valoración de información y aplicación de sus conocimientos, determinaría unos objetivos reales coincidentes con los previstos y movilizaría la actividad docente y discente hacia una didáctica que obligaría a los alumnos a pensar por sí mismos, a emitir y justificar juicios de valor, a resolver problemas con sentido para ellos, etcétera. En otras palabras, un tal sistema de exámenes sería una poderosa palanca para impulsar la calidad de la educación.

Dado que la evaluación del aprendizaje puede influir positiva o negativamente en la calidad de la educación, resulta necesario determinar qué características debe poseer un modelo evaluativo para maximizar sus efectos positivos y minimizar o anular los negativos.

Evidentemente, la condición fundamental de un sistema de evalua-

ción es que sea educativamente válido, lo que implica no solamente que los instrumentos de medida utilizados (tests, pruebas, escalas de observación, procedimientos no formales, etc.) sean válidos en el sentido psicométrico, sino también que exista un máximo de congruencia entre los criterios y modos de evaluación, la especificación de objetivos y el proceso didáctico o de intervención educativa [8], garantizando de este modo la funcionalidad, eficacia y eficiencia de la enseñanza y el aprendizaje, que en conjunto constituyen los criterios básicos de calidad de la educación.

### *La evaluación del rendimiento académico en la Universidad*

Determinar, identificar y categorizar los modos y características de la evaluación de los alumnos en la educación universitaria implica plantearse y, consecuentemente, contestar, al menos, tres cuestiones básicas:

- ¿Qué se pretende evaluar en la Universidad?
- ¿Cómo se evalúa lo que se pretende evaluar?
- ¿Para qué se evalúa?

La respuesta a la primera pregunta ¿qué se pretende evaluar?, significa identificar cómo conciben los profesores los resultados o producto de su actividad y del aprendizaje de los alumnos. Parece que existe unanimidad en la proposición de que el aprendizaje en la Universidad, el producto de la educación universitaria, es una realidad compleja y, en cierto grado, evasiva. En términos muy generales, podemos decir que los profesores universitarios coinciden en la noción de que el producto educativo se identifica con la adquisición de nuevos conocimientos, nuevos modelos de comportamiento intelectual y nuevas predisposiciones y actitudes hacia la ciencia y la cultura. Paradójicamente, sin embargo, un gran número de profesores, que consideran complejo el resultado o producto esperado de la educación, y de modo similar las metas y objetivos reales de la enseñanza universitaria, no parecen percatarse de la complejidad de su evaluación. Es decir, evalúan el aprendizaje de los estudiantes como si los resultados y objetivos de la enseñanza fueran simples, asignando notas y calificaciones con criterios que tienden a hipersimplificar la realidad evaluada.

Los modos prevalentes de evaluación en la Universidad ponen primariamente el énfasis en el conocimiento de contenidos. Puede aventurarse que esta situación es resultado de varios factores. De una parte, los profesores se han formado en Facultades organizadas en torno a disciplinas, en las que la atención se centra en la asimilación de la gran cantidad de información acumulada y organizada por otros profesores

y por ellos mismos a lo largo de años. El dominio de un modo de búsqueda e investigación, que constituye la esencia de una disciplina, queda relegado a un segundo plano. De otra, los profesores mismos se organizan en cátedras y departamentos a su vez divididos en subunidades que se corresponden con aspectos específicos de las disciplinas. La oferta docente de los departamentos se concreta en cursos escasamente relacionados entre sí, que cubren un núcleo determinado de conocimientos de cuyo dominio son examinados los estudiantes al finalizar el proceso. Las pruebas, exámenes, notas y calificaciones se hallan íntimamente vinculados a cada curso. La limitación de cada curso individual y la carencia de una real organización acumulativa y secuencial de los mismos impide de facto determinar productos o resultados educativos que trasciendan cada curso y disciplina entendida ésta como un conjunto de conocimientos-contenido. La evaluación en este contexto, por su parte, refuerza exclusivamente la cobertura de la materia incluida en el programa que, en general, tiene escasa relación con la comprensión y habilidad para aplicar ideas ni con la competencia para adquirir y usar el conocimiento, y que, teóricamente, la mayor parte de los profesores consideran resultados esperados de la enseñanza universitaria.

En síntesis, la evaluación del aprendizaje de los estudiantes universitarios se limita, en general, a la determinación del producto o resultado de la instrucción formal, entendiendo por tal la materia de una disciplina explícitamente enseñada. Esta situación supone un grave reduccionismo en la conceptualización práctica del producto de la educación cuya superación exige un serio esfuerzo teórico en el análisis de los objetivos deseables en las carreras y disciplinas universitarias y un gran esfuerzo de imaginación en la preparación de los programas evaluativos.

La respuesta a la segunda cuestión ¿cómo se evalúa lo que se pretende evaluar?, supone, al menos, la identificación de tres características de la evaluación universitaria: medios e instrumentos usados, criterios y referencias para la interpretación.

a) Los medios de evaluación usados en la Universidad son congruentes con el producto realmente evaluado. Se trata, en general, de distintas formas de prueba para determinar el dominio de contenidos disciplinares. Los llamados «exámenes» consisten en la aplicación de pruebas escritas y orales. Sólo en determinadas Facultades se utilizan pruebas de realización práctica (laboratorio y otras) y, aún en estos casos, su peso en la valoración final es muy relativo. Siguen predominando las pruebas escritas no estructuradas, de respuesta abierta (ensayo o preguntas cortas), basadas en el recuerdo, sobre las estructuradas, de respuesta cerrada, comúnmente conocidas como pruebas objetivas, basadas en el reconocimiento.

El problema del predominio de estas pruebas no radica en las características de las mismas, sino en su utilización mecánica e indiscriminada, sin referencia clara a los objetivos cuyo logro se pretende determinar, ni a los potenciales usos de los resultados de la evaluación, más allá de la mera calificación de los alumnos.

En todo caso, la evaluación universitaria raramente abandona la vía de la situación específica de prueba para circular por las otras dos grandes avenidas instrumentales: la observación sistemática del comportamiento intelectual de los estudiantes y el análisis de los trabajos no específicamente realizados como prueba. Hemos de reconocer, sin embargo, que la dificultad intrínseca de aplicar adecuadamente estos medios, unida a la estructura y organización de la Universidad, así como a la situación real de la enseñanza universitaria (grupos de docencia excesivamente numerosos, calendario y horarios escasamente racionalizados, etc.) no favorecen la diversificación de las actividades evaluativas.

b) El juicio de valor en que esencialmente consiste toda evaluación se justifica en función de la comparación de las características de la realidad evaluada (el aprendizaje de los alumnos, en nuestro caso) con criterios y standars exteriores al mismo. Si preguntamos al profesorado universitario por los criterios que utiliza en su evaluación, después de citar el más vago y general, es decir, el «dominio de la asignatura», concretaría su respuesta en términos como los siguientes:

- Extensión de los conocimientos.
- Organización, coherencia e integración de los mismos.
- Razonamiento lógico.
- Originalidad.
- Objetividad.
- Etc.

No siempre es tarea fácil, sin embargo, identificar los criterios de una calificación, por qué la realización de un alumno ha sido valorada de «notable» o de «aprobado». Y esto ocurre porque, en la práctica corriente, los profesores no evalúan con una lista de criterios explícitamente formulados, sino que, con gran frecuencia, éstos surgen en el propio proceso de valoración. Parece que los criterios emergen de la experiencia, como explicación de la evaluación.

Pero, ¿es realmente posible especificar de antemano todos los criterios para valorar la calidad del aprendizaje de los alumnos? Si se acepta el principio de emergencia, la respuesta habrá de ser negativa, ya que aún partiendo de algunos criterios básicos, ninguna lista puede ser realmente exhaustiva. En efecto, existe siempre la posibilidad de que otros criterios surjan al enjuiciar el aprendizaje. No obstante, la imposibilidad de una especificación exhaustiva no justifica la renuncia a, o la misma

valoración de, la exigencia de precisar, hasta donde ello sea posible, las características que debe reunir el aprendizaje de los estudiantes, como base para su evaluación. Además, la explicitación de los criterios evaluativos permite su comunicación a los alumnos antes de comenzar el aprendizaje, facilitando así la función orientadora de la evaluación y actualizando su capacidad de influir en la calidad de la educación.

A la no especificación de los criterios, como rasgo altamente generalizado en la evaluación universitaria, hay que añadir la relativa complejidad de su utilización, como factor perturbador de la situación. En primer lugar, el número considerable de los que pueden ser empleados plantea el problema de su armonización. En segundo lugar, los criterios pueden aplicarse de forma distinta. Aun suponiendo que se dispusiera de un conjunto fijo de criterios, cada profesor los utilizaría de modo diferente, en función de su propia política evaluativa. Algunos profesores fijan niveles mínimos para cada criterio y el alumno es negativamente evaluado si no logra alcanzar el mínimo en alguno de ellos. Otros aplican reglas compensatorias simples: el nivel alto en un criterio compensa el bajo en otro, pudiendo llegar a definir una «función de ponderación», como intentan los esquemas analíticos de puntuación. En estos casos subyace implícita la idea de que la evaluación se identifica fundamentalmente con la valoración global, es decir, con la puntuación y la calificación en una materia. En tercer lugar, hay situaciones en que las reglas, incluso las mejores, no son aplicables sin distorsionar seriamente su función. Es decir, los metacriterios (los criterios para aplicar los criterios) son dependientes del contexto. Finalmente, hay criterios que son difícilmente expresables [9] y, sin embargo, están vigentes como reglas tácitas compartidas por la comunidad educativa.

Otro problema frecuente en la evaluación del aprendizaje de los alumnos es la discrepancia entre los criterios utilizados por distintos profesores con grupos diferentes de la misma disciplina y curso, y la falta de coherencia que perciben los alumnos al pasar de un curso a otro en la misma disciplina, lo que se traduce en una auténtica «subversión» de la función pedagógica de la evaluación y en un acto de discriminación injustificado.

En resumen, existen razones pedagógicas y morales para el establecimiento previo de los criterios por los cuales se va a juzgar el aprendizaje de los alumnos. Algunos criterios emergen directamente de la propia definición de la tarea de enseñanza y aprendizaje y, obviamente, resulta de capital importancia una adecuada especificación de esta tarea. Esta explicitación criterial contribuye a promover un concepto de excelencia y calidad en el aprendizaje. No obstante, la explicitación de criterios no es suficiente. La experiencia evaluativa compartida por alumnos



y profesores en la cual los criterios van surgiendo como fruto de la interacción es también importante en esta línea [10].

c) La interpretación de las puntuaciones y calificaciones de un sujeto producto de la evaluación suele hacerse con una doble referencia. Cuando la interpretación se hace con relación a las puntuaciones o calificaciones obtenidas por el grupo a que el sujeto pertenece (población estudiantil, curso, etc.) se habla de referencia normativa. Cuando la interpretación se hace con relación a un patrón de realización preestablecido, se habla de referencia criterial.

Los profesores universitarios generalmente no explicitan la referencia evaluativa y, en todo caso, combinan en la práctica ambos tipos de referencia. En primer lugar, utilizan la referencia criterial al calificar a los estudiantes por el dominio que muestran de la materia. Sin embargo, la calificación varía según el rendimiento medio del grupo o curso (referencia normativa). Es frecuente la contrastación de que un profesor aprueba normalmente a un determinado porcentaje de sus alumnos con independencia de las diferencias de nivel entre los distintos grupos o cursos.

Parece evidente que la referencia criterial resulta de mayor coherencia pedagógica al destacar con mayor nitidez lo que el estudiante debe ser capaz de hacer. No obstante, ambos tipos de referencia tienen sentido en la evaluación universitaria para satisfacer las exigencias de las diversas funciones que cumple o puede cumplir.

La respuesta a la tercera cuestión (¿para qué se evalúa en la Universidad?) supone identificar los usos que profesores y alumnos hacen de la evaluación. Brown [11] diferencia entre dos usos principales de la evaluación: la toma de diversas decisiones y recoger información sobre el estudiante para comprenderle y ayudarle. Entre las decisiones más importantes tomadas con base en la evaluación destacan la calificación de los estudiantes, determinante de su carrera universitaria, la selección de estudiantes, la certificación, titulación y acreditación académicas y la revisión de la política universitaria.

El «feed-back» a profesores y alumnos se concreta en un flujo fiable de información sobre la eficacia y otras características del aprendizaje y la enseñanza, que sirva de base para su perfeccionamiento.

Hemos de reconocer que el uso predominante de la evaluación en la Universidad es la puntuación y calificación de los estudiantes en cada materia, curso, ciclo y carrera. Evaluación y calificación son identificadas frecuentemente por profesores y alumnos e, incluso, en la bibliografía sobre el tema, olvidando la gran función perfectiva de la evaluación. La calificación es necesaria y útil para seleccionar, predecir y certificar, pero es, en el mejor de los casos, neutral en la mejora del apren-

dizaje y la enseñanza. Su contribución a que los estudiantes comprendan la conexión entre su actividad discente y los criterios para valorar los resultados y a participar en el control inteligente del aprendizaje es limitada. En otras palabras, las calificaciones y puntuaciones son ajenas a la evaluación formativa. Circunstancialmente, además, se produce un abuso de las calificaciones. A veces se puntúa generosamente a los estudiantes sin una justificación seria, suponiendo que de este modo se reducirá la protesta y reclamación estudiantil sobre las notas, o que se asegurará un mayor número de estudiantes en una materia optativa en cursos subsiguientes. También ocurre en ocasiones que se aprueba a un alumno con la caritativa, pero equivocada, intención de no deteriorar su autoconcepto y autoimagen, lo que, a más largo plazo, resulta contraproducente.

En mi opinión, estos abusos en la calificación descansan en el falso presupuesto de que evaluar la realización de un estudiante es equivalente a juzgar a una persona en cuanto persona. Los estudiantes saben que no todo lo que produce un ser humano reputado como inteligente y honesto es bueno, ni malo todo lo que produce un individuo con reputación negativa.

En resumen, las prácticas evaluativas en la Universidad y las ideas no siempre explícitas, que sobre tales prácticas sustentan los profesores, configuran tres concepciones de la evaluación, de los alumnos, con moderado predominio de la primera:

1) Los exámenes y las tesis, como las formas prevalentes de la evaluación universitaria, hacen muy difícil, o quizá imposible, determinar cualquier producto educativo que trascienda los cursos y las disciplinas. Este tipo de evaluación, en consecuencia, es un reforzador de la enseñanza y el aprendizaje, como expresión de materia cubierta, características de nuestra educación universitaria.

2) Como alternativa, algunos profesores adoptan la postura de que los objetivos y los productos de la educación están más o menos implícitos en el proceso, y de que la evaluación de los más importantes es probablemente imposible. Su conclusión es que la evaluación resulta innecesaria, si el estudiante ha estado sometido o expuesto a un ambiente instructivo adecuado.

3) Una tercera posición es la de aquellos que discuten los objetivos y productos educativos en que se concreta la evaluación universitaria y sugieren la necesidad de trascenderlos para destacar el valor de las llamadas metas expresivas. Esto supone un contexto o marco de referencia evaluativo más flexible, parcialmente, pero no exclusivamente definido por lo que el alumno ha hecho, por qué lo ha hecho y los resultados específicos obtenidos. Se trataría de situaciones evaluativas en que los

alumnos pudieran expresar sus propias ideas y a sí mismos, liberados hasta donde ello sea posible, de una estructura impuesta por las expectativas de los evaluadores.

*Modelos evaluativos generales y su proyección  
en la valoración del rendimiento académico*

Las tres concepciones o formas de evaluación universitaria previamente citadas podían considerarse como otros tantos modelos evaluativos; pero, acontece que la expresión «modelos de evaluación», tal y como ha sido acuñada y tratada en la bibliografía sobre el tema, hace referencia a paradigmas evaluativos generales, surgidos como explicación de diferentes prácticas en la evaluación de programas educativos —no precisamente de la determinación del aprendizaje de los alumnos— y, en consecuencia, la tipología de modelos no responde a las formas de evaluación universitaria que hemos identificado. No obstante, por tratarse de modos de concebir determinada parcela de la evaluación educativa (la evaluación del rendimiento académico) deben ser susceptibles de clasificación en uno o varios modelos generales de evaluación.

Pero antes de proceder a tal clasificación, veamos cuáles son estos modelos generales y cuál es su significado y alcance.

Desde 1975 hasta hoy se han publicado, al menos, ocho tipologías de evaluación diferentes, que responden a otros tantos criterios de clasificación, y arrojan un total de 47 categorías. Guba [12] identifica los modelos de evaluación con los paradigmas científicos (positivismo lógico-fenomenológico-naturalístico); Gowin y Green [13] con la cuestión de valor que la evaluación plantea respecto a la realidad (valor intrínseco, valor instrumental, comparativo, optimizador, etc.); Glass y Ellett [14] con los tipos determinados por el contexto y la metodología usada (decisión, determinación de progreso, descripción, etc.); Talmage [15] con los principales tipos de la investigación (experimentación, descripción, input-output, etc.); Stufflebeam [16] con enfoques evaluativos en función de su orientación (política, valorativa, etc.); el Joint Committee on Evaluation Standards [17] con grupos de normas o criterios (utilidad, viabilidad, exactitud, ética); Gardner [18] con marcos de referencia determinados por las distintas interpretaciones dadas a los componentes de la evaluación (juicio de expertos, medida, congruencia entre realizaciones y objetivos, toma de decisiones, etc.); House [19] con lo que denomina concepciones básicas de la evaluación (análisis de sistemas, objetivos conductuales, toma de decisiones, quasi-legal, etc.).

El análisis de estas tipologías o conjunto de categorías revela que, aunque pueden ser útiles al señalar algunas distinciones y tendencias

en el campo de la evaluación, estamos lejos de lograr un mínimo consenso. La descripción de cada categoría dentro de las tipologías ofrece algunas ideas útiles, pero están lejos de una descripción explícita y precisa.

Por otra parte, dado que las tipologías utilizan diferentes criterios para llegar en algunas ocasiones a rótulos semejantes, pueden entenderse que cada una de las cuarenta y siete categorías es única y que los conjuntos pueden cruzarse para un análisis exhaustivo. Este ejercicio proporcionaría más de 450.000 casillas distintas. En consecuencia, si consideramos el número de tipologías, de categorías y de casillas resultantes de su cruzamiento podemos afirmar que nos hallamos en presencia de una inflación en la identificación de modelos evaluativos, pero muy lejos de un razonable uso de los mismos. El mayor problema con los esquemas de clasificación aludidos radica en su interpretación y aplicación. Las tipologías de modelos, que en conjunto son resultado de variar el énfasis en los propósitos, métodos y componentes del proceso evaluativo, en la práctica se convierten con frecuencia en fuente de confusión y conflicto entre posiciones distintas, como es el caso, entre otros, de la estéril polémica entre modelos cuantitativos y cualitativos [20].

Ante esta situación Stake [21] sostiene que los llamados modelos evaluativos son en realidad reflejo de creencias e ideologías. Al denominarlos con el término modelo se crea la expectativa de que pueden proporcionar la suficiente orientación y los procedimientos específicos para asegurar una evaluación adecuada en diferentes contextos. Y, sin embargo, la realidad no es esa. Resulta poco razonable suponer que cualquiera de los modelos evaluativos podría responder adecuadamente a las exigencias complejas incluso del más simple de los estudios evaluativos.

Las técnicas y métodos para llevar a cabo una evaluación son una función del problema, de los objetivos, de los valores inherentes al contexto del programa a evaluar, y del tiempo y recursos que pueden dedicarse a esta tarea, factores todos altamente variables de una a otra situación.

Los problemas en la utilización de los llamados modelos evaluativos derivan de la tendencia natural a considerarlos como instrumentos más poderosos de lo que en realidad son para ciertos fines y, en consecuencia, resultan disfuncionales. Frecuentemente son identificados con metodologías para realizar evaluaciones concretas en vez de asignarles su función propia de marcos de referencia dentro de los cuales caben muy diversos constructos y métodos más específicos. Lo que realmente proporcionan los modelos es una guía o heurística para pensar acerca de cómo puede desarrollarse una evaluación. En mi opinión, el papel o

función heurística de los modelos no ha sido siempre tomado en consideración.

Para precisar un poco más el significado de los modelos de evaluación resulta útil recurrir a la distinción de Hesse [22] entre *modelos en ciencia* y *modelos de ciencia*. Los modelos en ciencia (modelo atómico, modelo de la estructura cristalina, modelo de conflicto, por ejemplo) generan observaciones de fenómenos que pretenden confirmar o refutar el modelo sin ambigüedades. Intentan ser precisos y básicamente cuantitativos potenciando el desarrollo de técnicas de medida para describir los fenómenos de interés. Son, asimismo, específicos. Se refieren a un número seleccionado de fenómenos, evitando intencionalmente resumir o integrar las características de un acontecimiento o entidad. Finalmente, los modelos en ciencia se conciben como verificables. De ellos se derivan eventualmente hipótesis para comprobar su precisión y evidencia empírica acumulada para determinar su exactitud y utilidad.

Los modelos de ciencia (positivista, hipotético-deductivo, etc.) son construcciones teoréticas o filosóficas a la luz de las cuales se interpretan las observaciones. Para estos modelos no hay instancias fuera de ellos mismos que permitan determinar sin ambigüedad lo que existe o lo que es verdad. Estos modelos no son precisos, específicos y verificables. Normalmente, se refieren a una gran variedad de fenómenos resumiendo y agregando variaciones en amplios conjuntos o categorías de acontecimientos. Dada su carencia de precisión y especificidad, resulta difícil contrastar su exactitud y utilidad mediante hipótesis y acumulación de evidencia empírica.

Los modelos evaluativos hoy disponibles, parecen acercarse más a los modelos de ciencia que a los modelos en ciencia. En muchas ocasiones no rebasan las creencias de sus propios creadores. Por consiguiente, cuando tales modelos son tomados por los usuarios de la evaluación como si tuvieran las características de los modelos en ciencia, y se les adscribe niveles de precisión, especificidad y verificabilidad que no poseen, se produce la situación de confusión señalada.

¿Cómo utilizar, pues, los modelos de evaluación? Una de las limitaciones de los modelos es que imponen restricciones implícitas sobre las clases de hipótesis rivales o explicaciones alternativas que pueden plantearse. Los modelos dividen los programas a evaluar en categorías predeterminadas, que puede interpretarse fácil, y equivocadamente, como que todo lo necesario para comprender los efectos de un programa educativo está incluido en el modelo (en sus categorías, estadios y pasos que contiene). Propiamente usado, un modelo evaluativo es una heurística para organizar el propio pensamiento sobre la evaluación. Solamente en un sentido muy lato y vago puede servir como un algoritmo para llevar a cabo una evaluación [23].

Un modelo determinado, normalmente, no puede responder a las múltiples cuestiones derivadas de las diversas perspectivas que representan todos aquellos involucrados directa o indirectamente en un programa educativo.

*Interpretación de la evaluación del aprendizaje universitario  
en función de los modelos generales*

Dadas las características de la evaluación del aprendizaje de los alumnos en la Universidad, señaladas anteriormente (identificación del producto como contenido, predominio de la situación de prueba —exámenes— como medio generalizado, carencia de criterios explícitos, referencia mixta —normativo-criterial—, y fundamentalmente orientada a la calificación), resulta difícil determinar su correspondencia con los presupuestos de los modelos evaluativos generales, pensados, como afirmaba, más para evaluación de programas que para valorar el rendimiento académico de los estudiantes. En cierto modo, podemos encontrar apoyos para la explicación de la realidad evaluativa universitaria en varios de los modelos generales y ninguno de ellos puede dar cuenta completa de tal realidad.

En primer lugar, la evaluación del rendimiento en la Universidad representa un ejemplo del modelo general denominado «juicio profesional o de experto» que aparece en las tipologías de Gardner [24] y House [25]. En efecto, el profesor en su curso y como juez de tesis de licenciatura y doctorado utiliza en la evaluación criterios no explícitos que se suponen compartidos por todos los profesionales de la enseñanza.

En segundo lugar, puede rastrearse también cierta correspondencia con el modelo de evaluación como «determinación del logro de objetivos» (Gardner, 1977; Glass y Ellett, 1980; y House [26], aunque tales objetivos rara vez aparecen especificados y operacionalizados.

En tercer lugar, la evaluación universitaria representa en cierto grado al modelo de «medida» configurado por Ebel [27] y recogido en las tipologías de Gardner [28] y Stufflebeam [29]. En efecto, el uso de pruebas cuyos resultados cristalizan en un número es la esencia de este modelo y no es infrecuente su empleo en la Universidad.

Finalmente, las formas de evaluación del aprendizaje de los alumnos universitarios encajan también en el modelo evaluativo concebido como «toma informada de decisiones» identificado por Glass y Ellett [30], Stufflebeam [31], Gardner [32] y House [33]. No parece que puedan albergarse muchas dudas acerca del carácter decisonal de la evaluación universitaria cuya función básica, de facto, es facilitar y

justificar la decisión de paso o no paso de un alumno en una asignatura o en un grado (licenciatura, doctorado).

Del análisis de las correspondencias entre las formas de evaluación del aprendizaje de los estudiantes y los modelos evaluativos generales, surge la tentación de afirmar que la evaluación de los alumnos en las Universidades españolas configura un nuevo modelo evaluativo. Se trataría de un modelo híbrido o ecléctico producto de la combinación más o menos justificada de al menos cuatro modelos identificados. Según este modelo la evaluación universitaria sería fundamentalmente un juicio de valor expresado por expertos, en cierto modo apoyado en algún tipo de medida, acerca del logro de unos objetivos no siempre explícitos, pero generalmente identificados con el dominio de contenidos fácticos de las materias, como base para tomar decisiones sobre el proceso académico de los estudiantes.

Naturalmente que se trata de un pseudomodelo cuyo valor sobre los demás es su carácter descriptivo de una parcela específica de la evaluación educativa y quizá con menos potenciabilidad heurística para pensar sistemáticamente acerca de la evaluación educativa que cualquiera de los señalados.

Quizá la única utilidad de esta descripción se derive de la posibilidad de tomarlo como punto concreto de partida para la necesaria y urgente tarea de investigación empírica y reflexión crítica, por parte de la comunidad universitaria, acerca del sentido, significado, funciones y validez de una faceta esencial de su responsabilidad profesional: la evaluación del aprendizaje de los alumnos.

Dirección del autor: Arturo de la Orden, Facultad de Filosofía y Ciencias de la Educación, Departamento de Pedagogía Experimental, Edificio B, Universidad Complutense, 28003 Madrid.

#### NOTAS

- [1] ORDEN, A. de la (1969) La evaluación del rendimiento educativo y la calidad de la enseñanza, *Revista de Educación*, p. 206.  
ORDEN, A. de la (1981) Evaluación del aprendizaje y calidad de la educación, *La calidad de la Educación* (Madrid, C.S.I.C.).
- [2] CALONGI, L. (1978) *Valutazione* (Brescia, La Scuola).
- [3] MILLER, C. M. L. y PARLETT, M. (1974) *Up to the mark*, *Society for Research into Higher Education* (Surrey, Guilford).
- [4] DEARDEN, G. J. (1976) *Assesment in a electrical and electronic engineering course*.
- [5] MARTON, F. y SALJO, R. (1976) On qualitative difference on learning. 1. Outcome and process, *British Journal of Educational Psychology*, CP, pp. C-EE.
- [6] ELTON, L. R. B. y LANVILLARD, D. M. (1979) Trends in research on student learning, *Studies in Higher Education*, 4, pp. 87-182.

- [7] FUCHS, L. S. y OTROS (1984) The effects of frequent curriculum-based measurement and evaluation in pedagogy, student achievement and student awareness of learning, *American Educational Research Journal*, 21:2, pp. 449-460.
- [8] FELDHOUSEN y OTROS (1977) Is a lack of instructional validity contributing to the decline of achievement test scores?, en LIPSITZ, L. (ed.), *The test score decline: Meaning and issues* (New Jersey, Englewood Cliffs, Educational Technology Publications).
- [9] POLANYI, M. (1962) *Personal knowledge* (London, Routledge and Kegan Paul).
- [10] SADLER, D. R. (1983) Evaluation and the improvement of academic learning, *Journal of Higher Education*, 54:1, pp. 60-79.
- [11] BROWN, F. G. (1980) *Guidelines for test use: A commentary on the Standards for Educational and Psychological Tests* (Washington D. C., N.C.M.E.).
- [12] GUBA, E. G. (1978) *Toward a methodology of naturalistic inquiry in educational evaluation* (Los Angeles, California, Center for the Study of Evaluation).
- [13] GOWIN, D. B. y GREEN, T. F. (1980) Two philosophers view education, *Educational Evaluation and Policy Analysis*, 2, pp. 67-70.
- [14] GLASS, G. V. y ELLETT, F. S. (1980) Evaluation research, *Annual Review of Psychology*, 31, pp. 211-228.
- [15] TALMAGE, H. (1982) Evaluation of Programs, *Encyclopedia of Educational Research* (New York, AERA, McMillan).
- [16] STUFFLEBEAM, D. L. (1980) An analysis of alternative approach to evaluation, *Educational Evaluation and Policy Analysis*, 2, pp. 5-20.
- [17] JOINT COMMITTEES ON EVALUATION STANDARDS (1981) *Standards for Evaluation of Educational Programs, Projects, and Materials* (New York, McGraw-Hill).
- [18] GARDNER, D. E. (1977) Five Evaluation frameworks, *Journal of Higher Education*, 8:1, pp. 21-37.
- [19] HOUSE, E. R. (1980) *Evaluation with validity* (Beverly Hills, California, Sage).
- [20] WANG, M. C. y WALBERG, H. J. (1983) Evaluating educational programs: An integrative, causal-modeling approach, *Educational Evaluation and Policy Analysis*, 5:3, pp. 347-366.
- [21] STAKE, R. (1981) Persuasions not models, *Educational Evaluation and Policy Analysis*, 3:1, pp. 83-84.
- [22] HESSE, M. (1976) Models versus paradigms in the natural sciences, en L. COLLINS (ed.) *Use of models in the social sciences* (London, Tavistock Publ.).
- [23] BORICH, G. F. (1983) Evaluation models: A question of purpose, not terminology, *Educational Evaluation and Policy Analysis*, 5:1, pp. 61-64.
- [24] GARDNER, D. E. (1977), o.c.
- [25] HOUSE, E. R. (1980), o.c.
- [26] GARDNER, D. E. (1977), o.c.
- [27] EBEL, R. L. (1979), *Essentials of educational measurement* (New Jersey, Englewood Cliffs, Prentice-Hall).
- [28] GARDNER, D. E. (1977), o.c.
- [29] STUFFLEBEAM, D. L. (1980), o.c.
- [30] GLASS, G. V. y ELLETT, F. S. (1980), o.c.
- [31] STUFFLEBEAM, D. L. (1980), o.c.
- [32] GARDNER, D. E. (1977), o.c.
- [33] HOUSE, E. R. (1980), o.c.

#### BIBLIOGRAFIA

- BORICH, G. F. (1983) Evaluation models: A question of purpose, not terminology, *Educational Evaluation and Policy Analysis*, 5:1, pp. 61-64.
- BROWN, F. G. (1980) *Guidelines for test use: A commentary on the Standards for Educational and Psychological Tests* (Washington D.C., N.C.M.E.).
- CALONGI, L. (1978) *Valutazione* (Brescia, La Scuola).
- DEARDEN, G. J. (1976) *Assessment in an electrical and electronic engineering course*.
- EBEL, R. L. (1979) *Essentials of educational measurement* (N. Jersey, Englewood Cliffs, Prentice-Hall).



- ELTON, L. R. B. y LANVILLARD, D. M. (1979) Trends in research on student learning, *Studies in Higher Education*, vol. 4, pp. 87-182.
- FELDHUSEN y OTROS (1977) Is a lack of instructional validity contributing to the decline of achievement test scores?, en LIPSITZ, L. (Ed.) *The test score decline: Meaning and issues* (New Jersey, Englewood Cliffs, Educational Technology Publications).
- FUCHS, L. S. y OTROS (1984) The effects of frequent curriculum-based measurement and evaluation in pedagogy, student achievement and student awareness of learning, *American Educational Research Journal*, 21:2, pp. 449-460.
- GARDNER, D. E. (1977) Five evaluation frameworks, *Journal of Higher Education*, 81, pp. 21-37.
- GLASS, G. V. y ELLETT, F. S. (1980) Evaluation research, *Annual Review of Psychology*, 31, pp. 221-228.
- GOWIN, D. B. y GREEN, T. F. (1980) Two philosophers view education, *Educational Evaluation and Policy Analysis*, 2, pp. 67-70.
- GUBA, E. G. (1978) *Toward a methodology of naturalistic inquiry in educational evaluation* (Los Angeles, California, Center for the Study of Evaluation).
- HESSE, M. (1976) Models versus paradigms in the natural sciences, en L. COLLINS (Ed.) *Use of models in the social sciences* (London, Tavistock Publ.).
- HOUSE, E. R. (1980) *Evaluation with validity* (Beverly Hills, California, Sage).
- JOINT COMMITTEE ON EVALUATION STANDARDS (1981) *Standards for Evaluation of Educational Programs, Projects, and Materials* (New York, McGraw-Hill).
- MARTON, F. y SALJO, R. (1976) On qualitative differences on learning. 1. Outcome and process, *British Journal of Educational Psychology*, 46, pp. 4-11.
- MILLER, C. M. L. y PARLETT, M. (1974) *Up to the mark. Society for Research into Higher Education* (Surrey, Guilford).
- ORDEN, A. de la (1969) La evaluación del rendimiento educativo y la calidad de la enseñanza, *Revista de Educación*, 206.
- (1981) Evaluación del aprendizaje y calidad de la educación, *La calidad de la Educación* (Madrid, C.S.I.C.).
- POLANYI, M. (1962) *Personal knowledge* (London, Routledge and Keegan Paul).
- SADLER, D. R. (1983) Evaluation and the improvement of academic learning, *Journal of Higher Education*, 54:1, pp. 60-79.
- STAKE, R. (1981) Persuasions not models, *Educational Evaluation and Policy Analysis*, 3:1, pp. 83-84.
- STUFFLEBEAM, D. L. (1980) An analysis of alternative approach to evaluation, *Educational Evaluation and Policy Analysis*, 2, pp. 5-20.
- TALMAGE, H. (1982) *Evaluation of Programs. Encyclopedia of Educational Research* (New York, AERA, McMillan).
- WANG, M. C. y WALBERG, H. J. (1983) Evaluating educational programs: An integrative, causal-modeling approach, *Educational Evaluation and Policy Analysis*, 5:3, pp. 347-366.

SUMARIO: El autor resalta la función de la evaluación como elemento básico para mejorar la calidad del proceso enseñanza-aprendizaje y de su resultado. La evaluación del rendimiento académico en la Universidad supone una respuesta a tres interrogantes: ¿Qué se pretende evaluar en la Universidad? ¿Cómo se evalúa lo que se pretende evaluar? ¿Para qué se evalúa? Los distintos modelos evaluativos generales pueden servirnos de guía a la hora de responder a estas preguntas y desarrollar la tarea evaluativa.

Descriptores: Higher Education, Evaluation, Educational quality.