

# LA INVESTIGACIÓN SOBRE LA EVALUACIÓN EDUCATIVA

*por*  
*Arturo de la Orden*

## 1. Presupuestos teóricos

La evaluación constituye, sin duda, uno de los procesos nucleares en el contexto de la enseñanza, y de la educación en su conjunto. Formalmente constituye la tercera fase 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 tercera fase 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 relevante para describir la faceta evaluada (aprendizaje de los alumnos, eficacia del profesor, de los programas, de los métodos, de los centros, etc.). 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 heteroevaluación, auto o heteroreferencia, 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 la faceta educativa a evaluar.
- 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 promocionar a un profesor?, ¿para...?

Parece evidente que si la evaluación es un elemento o componente esencial del proceso educativo total, la concepción prevalente de la educación y las características de su realización determinarán en gran medida la naturaleza y 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 (De la Orden, 1969 y 1981) 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 a los alumnos en los exámenes, definen los objetivos reales del aprendizaje y la enseñanza. Calonghi (1978) 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, forma 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. Lo que se exige y como se exige en estas situaciones evaluadoras, se convierte pronto en los objetivos reales de la acción docente y educadora del profesor. La evidencia de este sentido es abrumadora. Los dos últimos años de la escuela secundaria americana adaptan sus programas para facilitar a los alumnos la superación de las pruebas nacionales estandarizadas (SAT - Scholastic Aptitude Test; y ACT - American College Tests) cuyos resultados se consideran necesarios para acceder a las universidades, especialmente a las más prestigiosas. En Inglaterra, los centros primarios comenzaban a preparar a los alumnos con dos años de anticipación para el llamado examen de los 11 años (eleven plus examination) que determinaba el tipo de escuela secundaria a que un alumno podía asistir; en la actualidad ocurre otro tanto con las pruebas finales de enseñanza media. En el contexto español, son conocidos los efectos sobre los objetivos de la enseñanza media de los antiguos exámenes de Estado y las re-

válidas, sin olvidar las pruebas de ingreso a las Escuelas Especiales que dieron lugar a la proliferación de las academias de preparación cuyas características son harto conocidas. Finalmente, no puede dejar de constatarse como las actuales pruebas de acceso a la Universidad 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.

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. Este es, a mi juicio, el significado profundo de la frase, frecuentemente empleada con sentido peyorativo y parcial, cuando se dice de un centro o de un profesor que «enseñan para los exámenes». En realidad, y dentro de ciertos límites, toda enseñanza y todo aprendizaje son orientados por el sistema de evaluación. Querámoslo o no, de forma consciente o inconsciente, la actividad educativa de alumnos y profesores está en algún grado canalizada por la evaluación del aprendizaje, por los exámenes, respondan o no a objetivos y programas previamente estipulados. Este es un hecho que, en sí mismo, no puede reputarse como positivo o negativo, pero cuyas consecuencias para la calidad de la educación son evidentes.

En efecto, si la calidad de la educación está vinculada a un sistema de coherencia 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— y, a través de ellos con los fines generales de la educación y el sistema de valores del que derivan, 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, 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 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, etc. En otras palabras, un tal sistema de exámenes sería una poderosa palanca para impulsar la calidad de la educación.

Aunque estos ejemplos son ficticios, para ningún profesional de la educación pasa desapercibido que pueden tomarse como representación, si bien esquemática, deformada y caricaturesca de situaciones reales. Aludiré aquí a una de ellas

por lo que pueda tener de estímulo para intensificar la investigación de este campo.

En un estudio realizado por un equipo bajo mi dirección (De la Orden, 1981b) sobre las características de las pruebas de 8.º curso de Educación General Básica que, por prescripción legislativa, deben elaborar los centros y aprobar la Inspección Técnica correspondiente, se constató, entre otros hallazgos, que más del 60% de los ítems de la subprueba de lengua española (se analizaron las subpruebas de lengua y matemáticas) estaban dirigidos a evaluar objetos clasificables en la primera categoría –Conocimiento– de la taxonomía de Bloom. Es decir, si se acepta nuestra tesis sobre la influencia de la evaluación, los exámenes que coronan la formación general, común a todos los españoles, definen unos objetivos reales de la enseñanza lingüística mayoritariamente de memorización de aspectos gramaticales, en abierta contraposición (incoherencia) con las metas establecidas en las Orientaciones Pedagógicas para la Educación General Básica (1971). Si la evaluación es tan poderosa como pensamos en el condicionamiento del proceso y el producto educativo, habremos de reflexionar seriamente sobre la calidad del aprendizaje de la lengua española en los centros educativos, al menos en la medida en que pueden generalizarse los resultados obtenidos en la muestra que sirvió de base a nuestro estudio.

Para la recogida de información necesaria para la descripción y expresión sintética de la situación en que se halla la faceta del aprendizaje a evaluar, el evaluador tiene a su alcance tres grandes vías de acceso:

- La primera es la observación sistemática del comportamiento del sujeto, utilizando para ello listas de control en las que figuran los aspectos a observar o las unidades que han de ser objeto de análisis e interpretación.
- La segunda es el análisis permanente del trabajo escolar realizado por los alumnos como fruto de su actividad ordinaria. Esta es una práctica firmemente establecida ya en los centros educativos, puesto que cada profesor normalmente corrige los ejercicios y cuadernos del estudiante con una cierta regularidad. El análisis que ahora proponemos implicaría un enfoque más crítico de esta labor correctora y el empleo de técnicas y procedimientos más objetivos y basados en criterios de mayor discriminación.
- La tercera vía son las pruebas de todo tipo, que hemos de concebir como trabajos que ha de realizar el sujeto expresamente diseñados con finalidad de control. Se trata de instrumentos diagnósticos cuyos resultados indican la suficiencia o insuficiencia en el proceso de alcanzar unos objetivos dados. Las pruebas que pueden ser utilizadas en educación son de carácter variado, desde un test normalizado, hasta una pregunta oralmente formulada en un momento oportuno, sin olvidar el empleo de toda la gama de pruebas objetivas, entrevistas, cuestionarios, exámenes orales y exámenes escritos de tipo ensayo. La elección de uno y otro tipo de prueba depende de la naturaleza y características del objetivo cuya consecución deseamos comprobar.

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 evaluación es

que sea educativamente válido, lo que implica no sólomente que los instrumentos de medida utilizados (tests, pruebas, escalas de observación, procedimientos no formales, etc.) sean válidos en el sentido tradicional, 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. Mager (1973) discute esta triple congruencia y señala que el elemento común entre criterios de evaluación, objetivos y proceso docente-discente es el comportamiento cognitivo, afectivo y psicomotor de los alumnos claramente especificado. En efecto, los objetivos no son otra cosa que la especificación de este comportamiento o de la capacidad para el mismo en forma de resultados esperados del proceso educativo; los criterios de evaluación, lo que se exige, constituyen, en conjunto, una muestra representativa de tal comportamiento; y el proceso didáctico se apoya en este comportamiento como base para su adquisición.

En consecuencia, la validez educativa de la evaluación del aprendizaje, como base de la calidad de la educación, significa:

- a) Que el aprendizaje de los alumnos (conocimientos y otros comportamientos que han de ser adquiridos) sea claramente especificado en forma de objetivos formulados sin ambigüedad. Para que los exámenes no se desvíen de los objetivos de enseñanza, es necesario conocer con la máxima precisión tales objetivos.
- b) Que los criterios de evaluación (lo que las pruebas exigen) deben constituir una adecuada muestra representativa de los contenidos y conductas especificados en los objetivos. Esta es la condición básica de la evaluación como palanca de la calidad de la educación. Si un objetivo básico de la enseñanza de la lengua española es que los alumnos expresen adecuadamente, oralmente y por escrito, sus propias ideas, los exámenes no deben concentrarse en exigir a los alumnos el reconocimiento o evocación de las reglas gramaticales, sino que deben solicitar de los estudiantes que expresen verbalmente su pensamiento de acuerdo con unos criterios, generalmente coincidentes con las normas gramaticales. Es de experiencia común, como afirmaba en el apartado anterior, que una de las formas más efectivas de asegurar que un objetivo sea puesto de relieve en la clase es evaluar periódicamente el grado en que se ha conseguido. Todo profesor ha comprobado que los estudiantes aprenden mucho más y de manera más efectiva todo aquello que es objeto de evaluación y control. La evidencia experimental ha confirmado claramente esta hipótesis. Basta citar el estudio comparativo de Walker y Schaffarzick (1974) sobre evaluaciones de curriculum en el que se puso de manifiesto que cuando las pruebas de rendimiento de los alumnos están en línea con las metas del programa innovador éste resulta más eficaz que el programa tradicional con el que se compara. Por el contrario, cuando las pruebas no cumplen esta condición, resultan superiores los programas tradicionales.

Por otra parte, el intento de alinear los criterios de evaluación con los objetivos contribuye decisivamente a clarificar a éstos y a resaltar su importancia en el pensamiento de profesores y alumnos. Las metas educacionales se hacen más definidas y más significativas, incrementando la coherencia entre criterios de evaluación, objetivos formales y aspectos del aprendizaje considerados por alumnos y profesores como resultados educativos valiosos e importantes.

- c) Que la forma, procedimientos e instrumentos de evaluación exijan los comportamientos especificados en los objetivos del modo más directo. Es decir, la técnica evaluativa empleada (observación de procesos, análisis de productos, informe, examen escrito tradicional, cuestiones de recuerdo o reconocimiento, pruebas con materiales o sin materiales, exámenes orales, etc.) debe estar directa y estrechamente relacionada con las características del aprendizaje o realización deseadas. Si un objetivo en la enseñanza de la Química es que los alumnos manejen adecuadamente determinadas piezas del equipo del laboratorio, exigir la descripción oral de cómo manejaría tales piezas, o la elaboración de un esquema gráfico del proceso, con formas de evaluación menos válidas que requerir el desarrollo real del proceso en el propio laboratorio.
- d) Que la evaluación sea fiable y objetiva en el sentido de que el azar o los errores instrumentales tengan un efecto mínimo en los resultados.
- e) Que el juicio valorativo acerca del aprendizaje de los alumnos sea expresado en términos de su adecuación a los objetivos más que en términos de la posición que ocupa su puntuación en el conjunto de las del grupo. Es decir, la evaluación debe reflejar lo que el alumno ha aprendido realmente más que su relación con lo que otros han logrado.

Feldhusen y otros (1977) aluden a otra condición de la evaluación educativa al tratar de argumentar en favor del concepto de «validez instructiva de los tests» como alternativa a la mera validez de contenido tradicionalmente contemplada. La validez instructiva asume la validez del contenido, pero requiere también a juicio de estos autores, la congruencia entre evaluación y proceso de instrucción. Por ejemplo, un examen de tipo ensayo para evaluar la capacidad de sintetizar ideas tendría una validez instructiva cuestionable si los alumnos no han tenido práctica alguna en organización de ideas o en redacción de ensayos.

Desde mi punto de vista, esta exigencia es más que una condición de validez de la evaluación, una consecuencia del cumplimiento de las cinco condiciones reseñadas. En efecto, supuestas tales características y comunicadas eficazmente a los estudiantes las expectativas del aprendizaje, es decir, el modelo de comportamiento (conocimientos y realizaciones) que ha de ser adquirido, aparecerá con mayor nitidez la naturaleza del proceso instructivo para alcanzarlo. La selección y organización del contenido, el tipo de experiencias discentes, el modo y la secuencia de interacción entre profesor, alumno, y materiales serán diseñados y realizados en función de su efectividad para lograr los objetivos previstos, única vía de satisfacer las exigencias de la evaluación.

Una evaluación así concebida supone, en principio, el requisito esencial para asegurar el sistema de coherencias en que consiste la calidad de la educación. La relación directa y la correspondencia entre fines de la educación, objetivos formales, objetivos reales, procesos instructivos y resultados de la educación, de tal manera que resulten isomórficos (Shoemaker, 1975), constituye la única garantía de un sistema con un alto grado de coherencia interna y externa con las expectativas y los resultados.

La cuestión, no obstante, sigue siendo ¿cómo operacionalizar esta concepción de la evaluación educativa y lograr formas de examen que garanticen el isomorfismo entre universos de objetivos educacionales y universos de criterios evaluativos, representados por universos de ítems, cuestiones y exigencias de evaluación?

Una vía posible y prometedora para operacionalizar el concepto de una evaluación isomórfica con los objetivos del aprendizaje, como base de la calidad de la educación, consiste, precisamente, en el diseño de procesos evaluativos inspirados en la referencia a criterios, más que a normas.

## 2. Objetivos de la línea del Departamento de Madrid

Partiendo de este supuesto, la línea de investigación iniciada en el Departamento de Pedagogía Experimental y Orientación de la Universidad Complutense pretende:

- Identificar y categorizar los modelos de evaluación de los alumnos utilizados en los distintos niveles del sistema educativo.
- Determinar la relevancia de los conocimientos, aptitudes y habilidades que se pretenden evaluar, en función de su grado de coherencia con las metas y objetivos explícita o implícitamente aceptados.
- Análisis y valoración de las técnicas utilizadas.
- Análisis psicométrico de las pruebas utilizadas.
- Identificación de las referencias evaluativas.
- Estudio de las características psicométricas de las pruebas de referencia criterial (validez, fiabilidad, puntuación crítica, etc.).
- Estudio experimental de los efectos del modelo evaluativo en la coherencia y eficacia de la enseñanza.

## 3. Proyecto de investigación del Departamento

En consecuencia, dentro de esta línea, se incluyen estudios que pueden diferir ampliamente no sólo en cuanto al objetivo específico sino también en cuanto a su tipología general, alcance y metodología (diseño, tipo e instrumentos de medida y análisis), si bien, todos coinciden en la meta de analizar los procesos evaluativos en educación como base de su perfeccionamiento, en cuanto instrumentos de capital importancia para optimizar el aprendizaje de los alumnos.

Simplificando, podríamos clasificar los proyectos de investigación en esta línea en las siguientes categorías:

### 3.1. *Investigación descriptiva de las características generales de los modelos de evaluación del alumno prevalentes en los distintos niveles de enseñanza.*

La metodología generalmente utilizada es la encuesta a muestras representativas de profesores y/o alumnos, mediante un amplio cuestionario cuyos ítems hacen referencia a los aspectos más importantes de la programación, realización de la evaluación y la utilización de sus resultados.

El análisis estadístico de los datos incluye:

- Cálculo de porcentajes y/o proporciones y su representación gráfica por ítem o alternativa para la muestra en su conjunto.
- Determinación de la significatividad de las diferencias en la distribución de frecuencias entre alternativas ( $\chi^2$ ).
- Cálculo de porcentajes y/o proporciones por grupos y su correspondiente representación gráfica cuando se utilizan variables clasificatorias (tipos de facultad, tipos de profesor, etc.; materias o áreas en E.G.B. y B.U.P., etc.).
- Determinación de la significatividad de las diferencias entre grupos en la

distribución de las frecuencias por ítem.

- Están en estudio, la aplicación del análisis de correspondencias y otras técnicas correlacionales avanzadas susceptibles de proporcionar alguna información sobre algún tipo de factor a partir de datos nominales.

3.2. *Investigaciones sobre la validez instructiva de los sistemas de evaluación* de los alumnos prevalentes en los distintos niveles de enseñanza. Se trata de identificar el concepto de producto educativo implícito en los instrumentos de evaluación y su relevancia en función de su coherencia con las metas explícita o implícitamente aceptadas y/o alguna taxonomía de objetivos educacionales derivada de la teoría de la educación inspiradora del sistema.

La metodología general supone:

- Conceptualización de los productos u objetivos de la enseñanza.
- Determinación de categorías de objetivos y/o resultados educativos.
- Recogida de instrumentos de evaluación de muestras representativas de centros y profesores (por áreas geográficas, niveles de enseñanza, disciplinas, etc.).
- Clasificación de los ítems en las categorías adoptadas (Juicio de expertos, etc.).
- El análisis puede comenzar con  $\chi^2$  para determinar la significación de las diferencias en la distribución de los ítems entre las diversas categorías y distribuciones teóricas determinadas como valiosas por expertos o derivadas de teorías.

3.3. *Investigaciones sobre la eficacia de programas de entrenamiento* acerca de la jerarquización de los resultados educativos y su expresión en los instrumentos de evaluación.

La metodología prevista se concreta en la técnica Q aplicada antes, durante y al final del entrenamiento. Como análisis principal se utilizará la correlación entre las clasificaciones de los profesores y la clasificación criterio. También podría recurrirse a análisis factorial y «cluster analysis».

3.4. *Análisis psicométricos de las pruebas utilizadas por los profesores en la evaluación del rendimiento de los alumnos: validez, fiabilidad, análisis de ítems.*

La metodología en este caso es la implicada en la psicometría clásica, incluido el análisis factorial de ítems.

Desarrollo de modelos de análisis para pruebas de referencia criterial.

3.5. *Investigaciones sobre los efectos del modelo evaluativo* utilizado sobre la coherencia y eficacia de la enseñanza (calidad educativa).

La metodología prevista es la implicada en el diseño experimental clásico (factorial) utilizando como modelos de análisis ANOVA, MANOVA y ANCOVA.

#### 4. Proyectos en marcha o terminados

En esta línea de trabajo se están realizando una serie de estudios concretos dentro y fuera del Departamento de Pedagogía Experimental y Orientación de la Universidad Complutense, que pasamos a describir:



#### 4.1. *Categoría A*

##### 4.1.1. *La evaluación del rendimiento académico de los alumnos en la Universidad.* (Equipo Departamental dirigido por Arturo de la Orden).

El estudio pretende contestar a las siguientes cuestiones:

- ¿Cuáles son los modelos de evaluación de los alumnos en la Universidad?
- ¿Existen relaciones detectables entre los modos de evaluación y de enseñanza en la Universidad?
- ¿Puede afirmarse que los modos de evaluación se diversifican según el tipo de centro universitario (facultades humanísticas, científico-experimentales, etc.)?
- ¿Puede afirmarse que los modos de evaluación se diversifican según el tipo de profesor (numerario, no numerario, novel, con varios años de experiencia)?
- ¿Qué modificaciones sería razonable introducir en los modos de evaluación que determinarían a su vez cambios perfectivos en las formas de trabajo docente y discente y, en última instancia, en el rendimiento y calidad de la enseñanza universitaria?

La investigación en curso pretende, pues, determinar, categorizar y valorar los modelos de evaluación utilizados en la enseñanza universitaria española. El conocimiento y clasificación de estos modelos permitirá explicar, a su vez, las formas de enseñanza y aprendizaje prevalentes en la Universidad y permitirá sacar conclusiones y formular propuestas para perfeccionar tanto la evaluación como la propia enseñanza universitaria.

Asimismo, el estudio permitirá comprobar diversas hipótesis sobre la influencia de determinadas variables en la modalidad de evaluación prevalente, según tipo de Facultades y Centros (Ciencias, Ciencias Sociales, Humanidades, Ciencias Médicas, etc., por una parte, y por otra, Facultades profesionales y no profesionales) y según el tipo de profesorado (numerario, no numerario, etc.).

La hipótesis general es que en la Universidad española prevalece un modelo implícito de evaluación de los alumnos, no claramente explicado y justificado que, a su vez, responde a una concepción de la enseñanza superior que el profesorado no se cuestiona por considerarle un supuesto generalmente aceptado.

Dado que los diversos centros que integran una Universidad se especifican por el cultivo y la enseñanza de una especialidad científica, técnica, artística o profesional, el estudio tratará de contrastar la hipótesis de que el tipo de centro influye en el modo de evaluación de los alumnos, matizando el modelo general aludido en la hipótesis primera.

Asimismo, se pretende contrastar la hipótesis de que ciertas variables que se identifican con características del profesor (numerario, no numerario; y experiencia docente fundamentalmente) influye, a su vez, en el modo de evaluar a los alumnos.

El siguiente cuadro representa los submodelos de evaluación universitaria implicados en las hipótesis precedentes:

		Tipos de profesor			
		Numerario		No numerario	
		Novel	Con experien.	Novel	Con experien.
Tipos de Centros	Científicos				
	Artísticos				
	Humanidades				
	C. Sociales				
	Sanitarios				
	Técnicos				

Finalmente, se pretende también contrastar la hipótesis de que existen diferencias en la evaluación entre los primeros y los últimos cursos de carrera.

Los tipos de datos a recoger serán fundamentalmente:

- Formas de examen
- Tipos de pruebas
- Valoración que se concede a cada tipo de prueba
- Periodicidad de exámenes y valor de los exámenes parciales respecto a la calificación final.
- Modos de calificación
- Amplitud de las escalas de puntuación
- Información que se proporciona a los alumnos sobre la evaluación al comenzar el curso o trimestre.
- Porcentaje de alumnos que obtienen cada nivel de calificación.

La población se circunscribe a la totalidad del profesorado de las tres Universidades de Madrid (Complutense, Autónoma y Politécnica) excluyendo la UNED, que por sus características no es homologable en prácticas evaluativas a los centros de enseñanza presencial. La muestra estará constituida por 400 profesores seleccionados aleatoriamente por estratos (tipos de centros y tipos de profesores).

El instrumento principal de recogida de datos es un cuestionario en el que la mayor parte de las preguntas son de respuesta cerrada o de respuesta corta y objetiva, sobre los distintos aspectos incluidos en los tipos de datos. Se incluirán también en el cuestionario algunas preguntas abiertas con la intención de detectar las concepciones generales de los profesores acerca de la evaluación y la enseñanza en la Universidad.

El cuestionario construido consta de tres secciones fundamentales:

- a) *Programación de la evaluación.* En este apartado se realizan las preguntas sobre el *cómo, cuándo, con qué criterios, qué elementos, qué niveles de exigencia, qué medios*, etc., van a ser utilizados para la realización de la evaluación.

- b) *Realización de la evaluación.* En esta sección del cuestionario, las preguntas al profesorado son ya sobre el momento de la evaluación, es decir, sobre *qué criterios de calificación* se consideran más relevantes, sobre los *mínimos exigidos* para superr la evaluación, sobre los contenidos propios de la evaluación, sobre *quién* toma decisiones evaluativas, sobre el *tipo de evaluación concreta* que es utilizada, etc.
- c) *Utilización de los resultados de la evaluación.* Esta última sección indaga sobre el uso de los resultados de la evrollo de la evaluación, sobre la *detección de deficiencias evaluativas*, sobre las *causas del suspenso* del alumno, sobre las *tareas de retroalimentación al alumno*, para que pueda superar las evaluaciones negativas, etc.

Es de hacer notar que este cuestionario, cuidadosamente elaborado, se interesa fundamentalmente por conocer *hechos*, es decir, la realidad propia de la evaluación en la Universidad y no tanto el deseo o la idea que de la evaluación tiene el docente.

El cuestionario, aplicado a la totalidad de la muestra, será complementado con una entrevista personal a 24 profesores a fin de profundizar en los aspectos más importantes de sus concepciones y prácticas evaluativas.

El formato del cuestionario permitirá el tratamiento mecanizado de los datos que serán analizados estadísticamente de forma que los resultados, previa confirmación de su significatividad, permitan la identificación de tendencias y submodelos evaluativos que se presentarán gráficamente en forma de perfiles y las comparaciones previstas en las hipótesis.

Algunos datos del cuestionario y los recogidos en la entrevista serán objeto de un análisis cualitativo complementario del anterior.

El estudio finalizará con un conjunto de conclusiones acerca de las características de la evaluación en la Universidad y sus relaciones con los métodos de enseñanza, de forma que sirva de base a propuestas de innovaciones didácticas.

#### 4.1.2. *La evaluación del rendimiento académico de los alumnos en E.G.B. y B.U.P.* (J. Mafokozi. Tesis doctoral, dirigida por Arturo de la Orden).

La investigación responde a objetivos similares (con pequeñas diferencias) a los ya descritos para el estudio de la evaluación universitaria y supone también la elaboración de un cuestionario que deberá ser cumplimentado por profesores de E.G.B. y B.U.P.

### 4.2. *Categoría B*

#### 4.2.1. *Estudio de las pruebas finales de Educación General Básica.* (Arturo de la Orden y Antonio González Soler).<sup>1</sup>

A través del análisis de las pruebas se pretende un doble objetivo: por un lado, aportar evidencia sobre las características de la evaluación educativa del sistema escolar, por otro, determinar la cantidad y calidad de lo que en realidad ofrece la Educación General a nuestros alumnos.

A efectos de análisis se distinguen dos grandes apartados:

- a) Contenido de las pruebas
  - Adecuación de los contenidos

1. Estudio finalizado. Publicado en la Revista Española de Pedagogía, n.º 156, Abril-Junio, 1982.

- Nivel de los contenidos
- Capacidades (Bloom) que pretenden medir y peso relativo de cada capacidad en las pruebas.

b) Características formales

- Extensión de la prueba
- Tipo de pruebas (objetivas, ensayo, etc.)
- Adecuación del tipo de ítem a la capacidad que pretende medir.
- Adecuación de construcción de la prueba

### *Hipótesis y variables*

Por tratarse de un estudio exploratorio es muy aventurado establecer hipótesis. En realidad lo que se pretende con esta investigación es analizar un hecho real, las pruebas de 8.º de EGB en Lengua y Matemáticas, para poder generar hipótesis cuya comprobación podría ser objeto de posteriores estudios.

De cualquier manera, nuestro conocimiento de la ciencia pedagógica y nuestra experiencia con la realidad educativa nos inclina a pensar en la posible influencia de ciertas variables independientes sobre alguna de las variables dependientes.

A tal efecto, y de entre la numerosa lista de posibles variables independientes, se han seleccionado las siguientes:

- 1) Tipos de Centro: Estatal, Privado (ambos de más de 8 unidades).
- 2) Medio ambiente: Rural, Urbano
- 3) Extracción social del alumnado (aplicable sólo al medio urbano): Alta, Media, Baja.

Las variables dependientes seleccionadas son las siguientes:

- 1) Adecuación y nivel de los contenidos
- 2) Peso relativo de las capacidades
- 3) Adecuación de la construcción de las pruebas

En consecuencia, se han generado las siguientes hipótesis nulas:

- 1) No existen diferencias significativas con respecto a las tres variables dependientes arriba mencionadas que puedan ser atribuidas al tipo de Colegio (Nacional o Privado).
- 2) No existen diferencias significativas con respecto a dichas variables que puedan ser atribuidas a la extracción social del alumnado que atiende los Colegios (media, baja).
- 3) No existe efecto de interacción significativo entre las variables tipo de Colegio y extracción social del alumnado con respecto a las tres variables dependientes.
- 4) No existen diferencias significativas con respecto a dichas variables dependientes que puedan ser atribuidas al medio ambiente (rural o urbano) en que se hallan enclavados los Centros Nacionales.
- 5) No existen diferencias significativas atribuibles a la extracción social del alumnado (alta, media, baja) de Centros Privados de Madrid.

Se trata de una investigación *ex post facto*. Por tanto, no es posible la manipulación de las variables intervinientes ajenas al interés de los investigadores.

### *Selección de la muestra*

Se partió de las siguientes variables, como factores hipotéticamente relacionados con el tipo y calidad de las pruebas:

- Medio socio-económico del que proceden los alumnos. En esta variable se identificaron tres niveles: bajo, medio, alto.
- Tipo de Centro según la localidad donde está enclavado. Se identificaron dos niveles: rural y urbano.
- Tipo de Centro según el régimen jurídico. Se identificaron dos niveles: privados, estatales.

Con estos condicionantes se extrajo la muestra por el procedimiento de muestreo estratificado, identificándose los siguientes estratos:

- Urbano - estatal - clase media
- Urbano - estatal - clase baja
- Urbano - privado - clase alta
- Urbano - privado - clase media
- Urbano - privado - clase baja
- Rural - estatal

Como se puede observar, no se han incluido los estratos siguientes:

- Urbano - estatal - clase alta, por considerar que el más alto nivel socio-económico de los alumnos que asisten a Colegios estatales no es comparable a dicho nivel de los Colegios privados.
- Rural - privado, en todos sus niveles, por ser un número insignificante.
- Tampoco se tuvo en consideración el nivel socio-económico en los Colegios estatales rurales, ya que al tratarse de localidades de escasa población difícilmente se produce tal estratificación, bien por la existencia de un solo Centro o por el escaso número de alumnos de clase social alta, los cuales en su gran mayoría son enviados a centros privados en régimen de internado.

#### *Carácter de los datos y técnicas de obtención de los mismos.*

*1.ª variable dependiente: Adecuación de los contenidos.* Los datos relativos a esta variable consistirán en las puntuaciones asignadas a cada prueba como consecuencia de la aplicación de una escala expresamente construida y validada a tal efecto.

*2.ª variable dependiente: Peso relativo de las capacidades.* Los datos consistirán en la proporción de ítems destinados a la medida de cada capacidad con respecto al número de ítems de cada prueba.

*3.ª variable dependiente: Adecuación de construcción de la prueba.* Los datos consistirán en las puntuaciones asignadas a cada prueba como consecuencia de la aplicación de una escala construida al efecto.

#### *Criterios para la construcción de las escalas y recogida de datos.*

*1.ª variable.* El criterio vendrá proporcionado por los Cuestionarios Oficiales en Matemáticas y Lengua, interpretados por expertos del equipo investigador.

*2.ª variable.* El criterio utilizado lo proporciona la Taxonomía de Objetivos de la Educación (Dominio Cognoscitivo) de B. Bloom.

*3.ª variable.* El criterio vendrá dado por las normas de construcción de instrumentos de evaluación de los manuales de psicometría comúnmente utilizados (Cronbach, Thorndike).

#### *Análisis estadístico*

Dado el carácter de los datos se utilizan para la comprobación de las hipótesis

tests paramétricos en todos los casos, con referencia a la primera variable dependiente.

*Hipótesis 1.<sup>a</sup>, 2.<sup>a</sup> y 3.<sup>a</sup>.* El análisis de varianza de varios factores (2 x 2) nos permite probar las tres primeras hipótesis. En efecto:

Tipo de Centro	Extracción Social	
	Media	Baja
Nacional	n = 15	n = 15
Privado	n = 15	n = 15

$$\text{Test estadístico: } F = \frac{V_e}{V_i}$$

Este tipo de análisis nos permite probar la hipótesis nula de la influencia del efecto principal de la variable Tipo de Centro (1.<sup>a</sup> hipótesis), de la variable Extracción Social (2.<sup>a</sup> hipótesis), así como de la interacción de ambas variables (3.<sup>a</sup> hipótesis). Hay que hacer las siguientes salvedades:

- 1.<sup>a</sup>) En el nivel *Nacionales* de la variable Tipo de Centro, no se incluyen los del medio rural porque no son comparables a los privados, los cuales se encuentran enclavados en medio urbano. Se harán análisis diferentes para comparar Colegios Nacionales de Medio rural y Colegios Nacionales de Medio urbano (4.<sup>a</sup> hipótesis).
- 2.<sup>a</sup>) En la variable *Extracción Social* no se incluye el nivel *Alta* por considerar que los alumnos procedentes de este nivel asisten fundamentalmente a Centros Privados (los que asisten a Centros Estatales no son significativos). Se hará un análisis específico para probar la influencia de esta variable con sus tres niveles (Alta, Media, Baja) pero afectando exclusivamente a Centros Privados (5.<sup>a</sup> hipótesis).

*Hipótesis 4.<sup>a</sup>* Para probar esta hipótesis nula, se utilizó el test más frecuente para probar la significación de diferencia de medias, es decir, el test *t*.

$$H_0 : \mu_{\text{rural}} = \mu_{\text{urbano}}$$

$$\sigma_1 \neq \sigma_2$$

$$t = \frac{\bar{X}_{\text{urbano}} - \bar{X}_r}{\sqrt{\frac{S^2_{\text{urbano}}}{n_1} + \frac{S^2_{\text{rural}}}{n_2}}}$$

*Hipótesis 5.<sup>a</sup>*

En este caso nos encontramos con la diferencia de medidas de tres muestras:

$$H_0 : \mu_{\text{alta}} = \mu_{\text{media}} = \mu_{\text{baja}}$$

El procedimiento estadístico utilizado será el análisis de varianza de un solo factor y el test empleado:

$$F = \frac{V_e}{V_i}$$

Para la segunda variable dependiente en todas las hipótesis se utiliza  $\chi^2$ .

Los siguientes cuadros muestran los resultados correspondientes a la segunda variable, es decir, el tipo de capacidades medidas en las pruebas.

*Datos obtenidos en la segunda variable: Peso relativo de las capacidades medidas en las pruebas.*

*Pruebas de Lengua – 8.º de E.G.B.*

Centros Privados	<i>Capacidades medidas (% de ítems dedicados a cada una)</i>					
	Conoci- miento	Com- pren.	Aplica- ción	Análi- sis	Sínte- sis	Evalua- ción
Extracción Social Alta	52	15	8	23		2
Extracción Social Media	63	11	9	16		1
Extracción Social Baja	58,5	18	8,5	14,5		0,5
Total Centros Privados	15	9	18		1	
Centros Públicos						
-Urbanos:						
Extracción Social Media	58	17	10	14		1
Extracción Social Baja	44	16	17	21	0,5	1,5
-Rurales:	56	12	13	17		1
Total Centros Públicos urbanos	51	16	14	18	0	1
Total Centros Públicos	54	14	13	18	0	1
Total Centros Públicos y Privados	55	14	12	18	0	1

*Pruebas de Matemáticas – 8.º de EGB*

Centros Privados	Capacidades medidas (% de items dedicados a cada una)					
	Conoci- miento	Com- pren. .	Aplica- ción	Análi- sis	Sínte- sis	Evalua- ción.
Extracción Social Alta	52	33	14		0,5	
Extracción Social Media	72	16	12			
Extracción Social Baja	61	18	21			
Total Centros Privados	62	22	16		0	
Centros Públicos						
-Urbanos:						
Extracción Social M.	76	11	13			
Extracción Social B.	66	13	21			
-Rurales:	62	21	17			
Total Centros Públicos Urbanos	712	12	17			
Total Centros Públicos	67	16	17			
Total Centros Públicos y Privados	65	19	16		0	

*Análisis realizados en la segunda variable: Peso relativo de las capacidades medidas en las pruebas.*

*- Resultados de las pruebas de  $\chi^2$  para comprobar la significación estadística de las discrepancias entre las frecuencias reales y las frecuencias esperadas de items en cada una de las categorías de la Taxonomía de Bloom.*

*Pruebas de Lengua*

	$\chi^2$	g.l.
Todos los Centros	1.909,8***	5
Centros Urbanos: Públicos y Privados	9,09	5
Centros Urbanos: Clase media y baja	8,51	5
Centros Urbanos Privados: Clase media y baja	4,7	5
Centros Urbanos Públicos: Clase media y baja	11,66*	5
Centros Urbanos de clase media: Públ. y Priva.	10,98	5
Centros Urbanos de clase baja: Públ. y Priva.	13,61*	5
Centros Públicos: Urbanos y Rurales	5,32	5
Centros Privados: Clase alta, media y baja	12,13	10



*Pruebas de Matemáticas*

	$\chi^2$	g.l.
Todos los Centros	2.544,4***	5
Centros Urbanos: Públicos y Privados	5,05	5
Centros Urbanos: Clase media y baja	13,84	5
Centros Urbanos Privados: Clase media y baja	7,46	5
Centros Urbanos Públicos: Clase media y baja	6,43	5
Centros Urbanos Clase media: Públi. y Priva.	2,11	5
Centros Urbanos Clase baja: Públi. y Priva.	2,18	5
Centros Públicos: Urbanos y Rurales	12,55*	5
Centros Privados: Clase alta, media, baja	25,52**	10

\* p &lt; 0,05

\*\* p &lt; 0,001

\*\*\* p &lt; 0,001

4.3. *Categoría C*

No se han iniciado proyectos de este tipo hasta la fecha.

4.4. *Categoría D*4.4.1. *Análisis de ítems en las pruebas de evaluación del rendimiento en las asignaturas de Pedagogía Experimental.*<sup>1</sup> (F. J. Tejedor).

Se pretende revisar las estrategias para el análisis de los ítems incluidos por los profesores del Departamento de Pedagogía Experimental en las pruebas trimestrales y finales. El estudio ofrece una estrategia integral para el análisis de ítems en las pruebas internas de evaluación normativa. Al mismo tiempo, se plantea el problema del análisis de los ítems en pruebas de referencia criterial.

La metodología se concentró en un análisis factorial de los ítems.

4.4.2. *La enseñanza universitaria vista por los alumnos: Un estudio para la evaluación de los profesores en la enseñanza superior.* (J.J. Aparicio, R. San Martín y F.J. Tejedor).<sup>2</sup>

Desde nuestra perspectiva el valor del estudio radica en el estudio psicométrico de un cuestionario para la evaluación del profesorado universitario por parte de los alumnos (fiabilidad y validez).

La metodología básica se concreta en la correlación de cada ítem con el total, la correlación múltiple de cada ítem con los demás, la correlación de cada ítem con el n.º 59, tomado como criterio y el análisis factorial de los ítems que permite detectar ocho factores claramente interpretables.

4.4.3. *Construcción de una batería de pruebas para la medición de objetivos de desarrollo en los niveles de EGB (2.ª etapa) y BUP.* (R. Pérez Juste y un equipo del Departamento de Pedagogía Experimental de la UNED).  
*Objetivos*

1. Disponer de un instrumento objetivo para la medición de los objetivos de desarrollo, aplicable a:

1. Será publicado por la Editorial de la Universidad Complutense.

2. Publicado en Cuadernos de capacitación docente. D.E.I., Madrid 1982.

- 1.1. Validación de constructo de la teoría de aprendizaje del profesor García Hoz.
- 1.2. Diagnóstico del grado de desarrollo alcanzado en la formación intelectual por alumnos de diferentes niveles escolares.
- 1.3. Pronóstico del éxito en BUP. Más adelante, del éxito en diversas carreras universitarias de grado medio y superior.
- 1.4. Orientación de los escolares.
2. Contar con un instrumento que permita poner a prueba la existencia o no de diferencias en la formación intelectual en función de variables tales como:
  - 2.1. Sexo
  - 2.2. Niveles de edad
  - 2.3. Clases sociales
  - 2.4. Residencia urbana o rural
  - 2.5. Métodos, técnicas o procedimientos de enseñanza y/o aprendizaje.
  - 2.6. Técnicas de estudio, etc.

#### *Características*

- Cada uno de los tests de la materia se ceñirá al esquema básico de aprendizaje humano, quedando integradas las fases del mismo: receptiva, reflexiva, extensiva, adquisitiva, expresiva y práctica.
  - Los subtests de cada una de las fases estarán integrados por los items que pretendan medir las grandes funciones de las mismas: observación, relación, reconocimiento, comparación, análisis, resolución, síntesis, representación, expresión, etc.
  - Todas las pruebas tomarán como contenidos sobre los que medir el logro de los objetivos de desarrollo las materias escolares, sean estas asignaturas o áreas del saber.
  - Habrá tests para todas las principales materias y áreas: lenguaje, matemática, ciencias naturales, ciencias sociales...
  - Se pretende alcanzar una prueba final con los mejores items de las pruebas construidas sobre cada área del saber.
- 4.4.4. Dentro de esta categoría pueden inclinarse las partes dedicadas a construcción, validación y baremación de instrumentos de evaluación del rendimiento académico en los siguientes proyectos de investigación.
- 4.4.4.1. *La eficiencia lectora en los alumnos de 2.ª etapa de EGB de Verdún-Roquetas (Barcelona).*  
Tesis elaborada por María Antonia Cardona Hernández, bajo la dirección de Margarita Bartolomé. Departamento de Pedagogía Experimental, Universidad de Barcelona.
- 4.4.4.2. *Predicción diferencial de la comprensión lectora en niños de suburbio a partir del lenguaje oral y del vocabulario.*  
Tesis doctoral (en elaboración) de Julia Victoria Espin López, dirigida por M. Bartolomé, Departamento de Pedagogía Experimental, Universidad de Barcelona.

4.5. *Categoría E**La función didáctica de la evaluación*

Tesis doctoral (en fase de diseño) de Angel Medrano, dirigida por A. de la Orden. Departamento de Pedagogía Experimental, Universidad Complutense.

## NOTAS BIBLIOGRÁFICAS

- CALONGHI, L. (1978). «*Valutazione*». La Scuola, 4.ª edic., Brescia
- FULDHUSEN 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*». Educational Technology Publications, Englewood Cliffs, New Jersey.
- MAGER, R.F. (1974). *Measuring instructional intent*. Feavon, Belmont, California.
- ORDEN, A. de la (1969). «*La evaluación del rendimiento educativo y la calidad de la enseñanza*». *Revista de Educación*, n.º 206.
- ORDEN, A. de la (1981a). «*El control de la evaluación educativa. Un ejemplo*». *Boletín de la Inspección del Estado*, n.º monográfico 1 y 2.
- ORDEN, A. de la (1981b). *Evaluación del aprendizaje y calidad de la educación*. En «*La calidad de la educación*», C.S.I.C., Madrid.
- «Orientaciones Pedagógicas para Educación General Básica» (1971). *Revista Vida Escolar*, n.º 128-130.
- SHOEMAKER, D.M. (1975). *Toward a framework of achievement testing*. *Review of Educational Research*, VOL. 45, págs. 127-147.
- WALKER, D.F. y SCHAFFARZICK, J. (1974). *Comparing curricula*. *Review of Educational Research*, VOL. 44, págs. 83-111.

Arturo DE LA ORDEN HOZ  
D. Pedagogía Experimental y Orientación  
Facultad Filosofía y C. Educación  
Ciudad Universitaria  
MADRID - 3