
TRABAJOS DE INVESTIGACIÓN

INFLUENCIA DE LA EVALUACIÓN DEL APRENDIZAJE EN LA EFICACIA DE LA ENSEÑANZA

por
Arturo de la Orden
Universidad Complutense

RESUMEN

En la línea de documentar la ya constatada influencia de la evaluación educativa en el rendimiento escolar, este artículo ofrece los resultados más salientes de un estudio experimental realizado en catorce centros educativos de Madrid para identificar los rasgos generales y específicos de los procesos evaluativos en la educación básica.

Se contrastaron los efectos de cuatro modelos evaluativos (A, B1, B2 y C) sobre los resultados escolares de los alumnos de 8º curso de EGB en Lengua y Matemáticas, durante un año académico.

En estas materias resultaron superiores los efectos optimizantes de los modelos B1 (evaluación criterial y formativa) y A (evaluación continua y formativa).

El modelo evaluativo con menor efecto en el rendimiento es el C (evaluación sumativa, identificado como prevalente en la población de escuelas considerada).

ABSTRACT

This article presents the main results of a experimental study to identify the traits of the assessment processes linked to the teaching effectiveness.

The assessment models showing the higher effect on Language and Mathematics achievement of 8th grade students are those criterion-referenced, formative-oriented, and applied on a continuous schedule.

* Este artículo ofrece una parte de la Memoria del Proyecto subvencionado por el CIDE en la convocatoria de 1988 al equipo de investigación integrado por A. de la Orden, I. Asensio, R. Carballo, M. J. Fernández, A. Fuentes, J. M. García, J. L. Gaviria, A. Lázaro, E. López, J. Mafokozi y M. J. Martínez de Toda.

INTRODUCCIÓN

La evaluación constituye, sin duda, uno de los procesos nucleares en el contexto de la enseñanza y de la educación en su conjunto. En cuanto tal elemento o componente esencial del proceso educativo total, la concepción prevalente de la educación y las características de su realización determinan en gran medida la naturaleza y alcance del modelo evaluativo; pero, a su vez, el modelo de evaluación influye decisivamente en el proceso y producto de la educación y, en consecuencia, en su eficacia y calidad. (De la Orden, 1969, 1981, 1988; Crooks, 1988)

Si 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.

En esta perspectiva, la condición fundamental de un sistema de evaluación para optimizar el rendimiento educativo o, si se prefiere, para incrementar la eficacia docente e institucional es la *validez educativa*, lo que significa:

1) 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.

2) 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.

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.

3) 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, son formas de evaluación menos válidas que requerir el desarrollo real del proceso en el propio laboratorio.

4) 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.

5) 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 de 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 de aprendizaje, es decir, el modelo de comportamiento (conocimientos y competencias) que ha de ser adquirido, aparecerá con mayor nitidez la naturaleza del proceso instructivo para alcanzarlo. La selección y la 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?

OBJETIVOS

En consecuencia, el objetivo general del estudio se centra en la identificación de los rasgos generales y específicos de los procesos evaluativos en la educación básica que garanticen su función optimizante. Se trata, pues, de la búsqueda disciplinada de un modelo que cumpla las funciones formativa y sumativa de la evaluación, con referencias criteriosales y normativas adecuadas, apoyado en instrumentos fiables y válidos, viable en el contexto de nuestros centros docentes y que permita una fácil y fluida comunicación de los resultados educativos a los alumnos y a las familias.

En síntesis, la investigación intenta responder sobre la base de la evidencia empírica a las siguientes cuestiones:

1. ¿Cómo combinar las funciones formativa y sumativa de la evaluación?
2. ¿Cómo combinar las referencias criterial y normativa en la evaluación?
3. ¿Cómo combinar los diferentes modos, técnicas e instrumentos de recogida de información y medida en el proceso evaluativo para garantizar un máximo de funcionalidad?
4. ¿Cómo estructurar el proceso evaluativo en su conjunto en el contexto de la clase y el centro para asegurar su viabilidad y función optimizante?
5. ¿Cómo proceder a la calificación del alumno (juicio evaluativo) y a la comunicación de los resultados de la evaluación a los padres y a los propios alumnos?

Para ello se diseñaron y se aplicaron durante un año distintos modelos de evaluación a clases paralelas del mismo nivel o curso y asignatura, comparando los resultados educativos en el período que cubre la experiencia.

HIPÓTESIS GENERAL

El estudio, en consecuencia, intenta contrastar la siguiente hipótesis general: los distintos modelos de evaluación producen efectos diferentes sobre el rendimiento de los alumnos en las materias de Lengua Española y Matemáticas.

VARIABLES

La hipótesis incluye las siguientes variables:

A) Variables Dependientes

A.1) *Rendimiento académico*. Medido a través de pruebas objetivas «ad hoc» de referencia criterial construidas para medir el nivel de cumplimiento de los objetivos-contenidos de cada una de las dos materias (Lengua y Matemáticas) según aparecen formulados en los Programas Oficiales del M.E.C. y especificados para la investigación. Las pruebas de medición del rendimiento se construyen independientemente del modelo evaluativo aplicado.

A.2) *Actitud hacia la materia*. Serán medidas las actitudes hacia las dos materias básicas del currículum (Lengua y Matemáticas).

B) Variable tratamiento o variable independiente-experimental:

Modelos evaluativos. Se han definido cuatro modelos evaluativos bien diferenciados. Una primera fase del proyecto se ha centrado en definir los modelos de evaluación de manera que se consideren suficientemente diferenciados.

C) Variables de control

C.1) *Curso-edad*. Alumnos de 8º de E.G.B. Los modelos evaluativos se aplicarán a los grupos seleccionados en cada curso.

C.2) *Profesor* Variable importante a controlar. Se aplica cada modelo evaluativo, asignando entre 8 y 10 profesores a cada modelo.

C.3) *Tipo de centro*. (Privado-Público). Se seleccionó una muestra de colegios (privados y públicos) para realizar la experiencia.

C.4) *Variables aptitudinales*. Para asegurar su no incidencia o eliminar su posible influjo se midieron antes de la experiencia; fluidez verbal, comprensión lectora, factor verbal, factor numérico y estilo cognitivo (DIC).

DISEÑO DE INVESTIGACIÓN

Se trata de un *diseño experimental clásico* con pretensiones de control directo de las variables más importantes que pueden incidir en el desarrollo del mismo. Las posibilidades de generalización del mismo se incrementan por la inclusión de gran número de centros, profesores y alumnos participantes en el estudio.

La representación simbólica del mismo, para cada variable dependiente, sería la siguiente:

Cuadro 1
REPRESENTACIÓN DEL DISEÑO (PARA CADA VARIABLE DEPENDIENTE)

COVARIABLES	GRUPOS DE TRATAMIENTO	POSTEST
X1 X2 X3 X4 X5	TA (9 grupos)	Y1
	TB1 (8 grupos)	Y2
	TB2 (8 grupos)	Y3
	TC (10 grupos)	Y4

X1 = Fluidez verbal (PAL)

X2 = Factor verbal (ANA)

X3 = Factor numérico (SIM)

X4 = Comprensión lectora (CL)

X5 = Estilo cognitivo (GEFT)

TA = Grupos que siguieron el modelo evaluativo A

TB1 = Grupos que siguieron el modelo evaluativo B1

TB2 = Grupos que siguieron el modelo evaluativo B2

TC = Grupos que siguieron el modelo evaluativo C

Y1, Y2, Y3, Y4 = Medidas de rendimiento en Lengua española o Matemáticas

MUESTRA

Se intentó seleccionar una muestra con un número de centros suficiente para asegurar que cada modelo pudiera ser aplicado por varios profesores, que representaran en lo posible la gama completa de competencias y orientaciones docentes.

Los centros finalmente seleccionados y los modelos evaluativos asignados fueron los siguientes:

Cuadro 2
CENTROS SELECCIONADOS

CENTROS	Nº GRUPOS	PÚBLICO/PRIVADO	MODELO EV.
S. Juan Bosco	2	Pu	A
D. Juan Manuel	2	Pu	A
I. Véritas	2	Pr	A
Virgen de Europa	3	Pr	A
Méndez Núñez	2	Pu	B2
Aula Nueva	2	Pr	B1
Obispo Perelló	4	Pr	B1
Víctor Pradera	3	Pu	B2
Ntra. Sra. Sta. María	3	Pr	B2
Santo Domingo	2	Pr	B1
Zulema	2	Pu	C
Miguel de Unamuno	4	Pu	C
Liceo Versalles	3	Pr	C
Rab. Tagore	1	Pu	C

INSTRUMENTOS

Fueron seleccionadas las pruebas para medir ciertas variables presumiblemente relacionadas con la variable dependiente en orden a su control. En esta perspectiva se aplicaron:

1. Construir frases diferentes. (PAL)
2. Símbolos. (SIM)
3. Analogías. (ANA)
4. Prueba de comprensión lectora (CL)
5. Test de figuras enmascaradas (GEFT)

Por otra parte, como medida de la variable dependiente se prepararon, a partir de los objetivos señalados en los programas oficiales (Programas Renovados, 1982), especificados por un grupo de especialistas, pruebas para medir el rendimiento en Lengua y Matemáticas. (PLE) y (MAT) También se elaboraron instrumentos para medir las actitudes de los alumnos hacia estas materias. (ALE) y (AMA).

Aunque el plan de investigación suponía la determinación de conocimientos previos, no pudo aplicarse ninguna prueba antes de comenzar el tratamiento.

Descripción de las pruebas

1. *Construir frases diferentes.* (PAL)

Se trata de elaborar 5 frases diferentes con cada uno de los dos grupos de tres palabras ofrecidas.

2. *Símbolos.* (SIM)

Pertenece a la *Batería Factorial de la inteligencia AMD* García Yagüe y Palomino López (1976). Ed Miñón.

Son 20 ejercicios en los que se trata de averiguar el valor de un signo desconocido combinando los datos que se ofrecen.

3. *Analogías.* (ANA)

Batería Factorial de la inteligencia AMD García Yagüe y Palomino López (1976). Ed Miñón.

Son 20 frases en las cuales hay una o dos palabras en letras mayúsculas que deben ser sustituidas por otras escogidas de entre las cinco ofrecidas de manera que se mantenga el sentido de la frase.

4. *Prueba de comprensión lectora.* (CL) Lázaro Martínez (1989)

Se presentan 18 textos sobre los que se hacen 28 preguntas, bien de elección múltiple, bien de asociación de conceptos, presentados en dos columnas.

5. *Test de figuras enmascaradas.* (GEFT)

Oltman, Raskin y Witkin (1971) Ad. esp. Fernández Ballesteros y Macía. (1981) Ed. TEA.

Consta de tres secciones con 25 ejercicios en los cuales se trata de encontrar una forma simple cuando ha sido enmascarada dentro de una figura compleja.

6. *Prueba de lengua.* (PLE)

La prueba de Lengua para 8º de E.G.B. incluye 27 ítems de elección múltiple, distribuidos de manera que recojan objetivos mínimos de todos los dominios señalados para la asignatura:

Expresión oral.	2 ítems
Léxico.	3 ítems
Semántica.	4 ítems
Vocabulario.	2 ítems
Técnica literaria.	2 ítems
Expresión escrita.	3 ítems

Gramática.	7 ítems
Literatura española.	4 ítems

7. *Prueba de matemáticas (PMAT)*

Consta de 30 ítems de elección múltiple, que recogen prácticamente todos los objetivos mínimos especificados.

8 y 9. *Prueba de actitudes lengua y matemáticas. (ALE y AMA)*

Consta de 42 proposiciones, 22 referidas a matemáticas y 20 a lengua. De todas ellas, 10 tienen un marcado carácter de rechazo hacia la materia. Los alumnos deben optar en una escala desde el 1, si su desacuerdo es total, hasta el 5, si se identifican plenamente con lo afirmado.

DESCRIPCIÓN DEL TRATAMIENTO

Modelos evaluativos

El núcleo esencial de la investigación lo constituyen los modelos de evaluación diseñados para probar su influencia sobre el rendimiento académico. Las dificultades del diseño de tales modelos, enmarcados dentro de un contexto determinado y fuertemente condicionados por la organización educativa (número de alumnos por clase, exigencias legislativas, etc.) han sido patentes. La implementación de los modelos se facilita sustancialmente con la implicación y colaboración del equipo de profesores aplicadores de cada modelo. En todos los casos, se precisa seguir con rigor las prescripciones operativas definidas en cada modelo, a fin de evitar en las aplicaciones de los respectivos equipos de profesores situaciones diferenciales que pudieran interferir en el efecto de cada uno de ellos.

Los modelos evaluativos son fundamentalmente tres. A partir de ahora los identificaremos como A, B, y C, aunque el B se desdobló en dos submodelos B1 y B2. Cada uno de ellos está definido en sus elementos esenciales (frecuencia de la evaluación, tipos de instrumentos, criterios de calificación, información a los alumnos, etc.).

Modelo A

Este modelo trata de acentuar el carácter formativo de la evaluación, pudiéndose definir como un modelo de evaluación continua, donde los elementos «formativos» están permanentemente presentes a lo largo del proceso de enseñanza-aprendizaje.

La evaluación formativa se vincula directamente con la mejora de la calidad de los procesos de enseñanza-aprendizaje; de hecho, «constituye la base de la enseñanza correctiva (recuperación) y de la autocorrección de la acción del profesor y del proceso didáctico, cuando todavía la situación es reversible, esto es, mientras se

desarrolla normalmente la secuencia enseñanza-aprendizaje y se puede canalizar con éxito hacia los objetivos previstos» (De la Orden, 1982).

A continuación se analizan algunos parámetros evaluativos que vienen condicionados por las características intrínsecas del modelo y que inciden en la programación, realización y uso de los resultados de las actividades evaluativas.

1. *Frecuencia de evaluación.*

De acuerdo con la filosofía del modelo enunciado, vinculada directamente con la mejora de la calidad de los procesos de enseñanza-aprendizaje, intentando detectar deficiencias, errores, etc., cuando todavía la situación es reversible, parece claro que la estrategia del evaluador es recoger los resultados de pruebas empíricas sobre la eficacia de los componentes de la secuencia pedagógica. En este sentido, se tenderá a aplicar pequeños ejercicios/actividades evaluativas para cada una de las unidades de contenido. La periodicidad de estas sesiones girará en torno a una quincena, siempre entendida con flexibilidad en función de las características (extensión) de la unidad temática correspondiente y del ritmo de aprendizaje del grupo de alumnos. No se excluye la valoración diaria del trabajo escolar.

2. *Tipos de pruebas.*

Se realizarán pequeñas pruebas de referencia criterial, variando según contenido/objetivo y con la frecuencia señalada.

Previamente se fijará el tipo de pruebas a utilizar para evaluar cada objetivo/contenido (ítems de prueba objetiva, pregunta corta, tema, problemas, etc.).

Teniendo en cuenta que la unidad temática (tema o conjunto de temas) constituye en este modelo la base de cada sesión evaluativa, parece lógico pensar que dichas sesiones estarán definidas por pequeñas pruebas.

3. *Definición de indicadores (criterios) de calificación de cada prueba evaluativa.*

Para la calificación del logro de cada alumno se tendrán en cuenta especialmente las prescripciones relativas a la construcción de instrumentos de medida de referencia criterial. Ello conllevará la definición de una escala de calificación.

La valoración de cada elemento de prueba (ítem o cuestión) se especificará previamente a la aplicación de las pruebas, dependiendo de las características intrínsecas del contenido a evaluar. A modo de ejemplo, en el caso de una prueba objetiva, la asignación de puntuaciones resulta relativamente sencilla, mientras que en una prueba de ensayo la valoración crítica del alumno, por ejemplo, puede ser en algún caso muy importante (peso relativo elevado) y en otros casos no tener tanta importancia (peso relativo bajo).

4. *Orientaciones sobre la escala a utilizar en la calificación y ponderación de los distintos elementos evaluativos.*

El modelo diferencia entre objetivos obligatorios (mínimos) y optativos (máximos). En función de ello consideramos que la calificación de suficiente sólo debería

asignarse estableciendo un punto de corte suficientemente elevado entre los objetivos obligatorios. A título indicativo sugerimos un punto de corte situado en el cumplimiento del 70% de los objetivos obligatorios. Por tanto, el alumno que supere el mencionado 70% de los objetivos obligatorios será considerado apto (suficiente) en el período evaluado y, por tanto, liberará materia. El que no alcance tal punto de corte será considerado insuficiente, debiendo recuperar la materia no superada (se especificarán los objetivos más deficientes dando las orientaciones «formativas» pertinentes para facilitar su recuperación.)

Dentro de la categoría de apto (suficiente) podrá especificarse entre: Aprobado, Bien, Notable y Sobresaliente. Dichas categorías vendrían determinadas por el dominio adicional de:

- El 30% de los objetivos obligatorios restantes hasta completar el 100%.
- El porcentaje de objetivos optativos máximos alcanzados.
- Otras actividades y trabajos realizados por el alumno susceptibles de evaluación de proceso por parte del profesor.

5. Tipo de información que se da al alumno antes de la evaluación.

a) *Al Principio de curso*, presentación detallada del modelo evaluativo, valoración e implicaciones formativas, destacando el carácter de evaluación continua.

b) *Periodicamente*, se resaltarán en las explicaciones de clase la importancia del trabajo continuado y sus consecuencias evaluativas.

c) *Antes de la aplicación de cada pequeña prueba evaluativa*, se motivará en torno al carácter formativo de determinación de progresos, mejoras, deficiencias y limitaciones, especificando también la valoración de cada parte de la prueba (diferencias entre objetivos obligatorios y optativos) y sus consecuencias posteriores dirigidas al logro de los objetivos no superados (eventuales procesos de recuperación).

6. Uso de los datos evaluativos.

a) Información a los alumnos:

— Se informará «permanentemente» a los alumnos de sus progresos y deficiencias.

— Se indicarán actividades/ejercicios para recuperar los objetivos no superados (es conveniente que en la programación de la enseñanza cada objetivo vaya unido a actividades de profundización o de recuperación para el logro del mismo).

— *Feed-back inmediato*: Se informará a los alumnos del rendimiento obtenido en cada una de las pruebas evaluativas lo más rápidamente posible.

— Conviene ofrecer extensa información al alumno sobre progresos y deficiencias, como complemento de la calificación asignada.

b) Información a padres:

Independientemente de que las calificaciones obtenidas por los alumnos en los distintos períodos evaluativos se reflejen en el correspondiente boletín de información a los padres, el equipo de profesores que lleve a cabo este modelo evaluativo,

considerará la posibilidad de establecer otras medidas de información a los padres, especialmente para casos de alumnos con rendimiento insuficiente y/o insatisfactorio.

Modelo B

Aunque también de carácter formativo y criterial, el modelo B se especifica por la realización de 5 evaluaciones parciales formales a lo largo del curso, con una programación previa del equipo de profesores e investigadores referida a periodicidad, fecha y duración de las pruebas, entre otros aspectos. Estas, tendrán carácter liberatorio, debiendo recuperar los objetivos no superados a partir de las orientaciones y recomendaciones que el profesor realizará esencialmente en grupo lo antes posible, desde la realización de las pruebas. Se pretende que el feed-back tenga un carácter inmediato.

Los criterios de evaluación harán referencia al logro de los objetivos, determinando previamente los niveles de exigencia y la ponderación de cada una de las partes de la prueba. Es evidente la necesidad de que estos niveles y ponderación sean semejantes en los profesores que apliquen este modelo.

La aplicación de distintos tipos de pruebas de evaluación dará lugar, en este caso, a dos submodelos, B1 y B2 diferenciados únicamente en este elemento.

1. Tipo de pruebas.

B.1) Los profesores de este submodelo utilizan una combinación de prueba objetiva y de otra prueba (comentario de texto o redacción para Lengua; problemas para Matemáticas).

B.2) Los profesores del submodelo B2 utilizan una combinación de preguntas cortas o tema y ejercicios similares a los del libro de texto o a los planteados en clase.

Para ambos submodelos quedará explícito que estas pruebas han de ser escritas y de carácter teórico y práctico.

2. Definición de indicadores (criterios) de calificación de cada prueba evaluativa.

En la programación de la evaluación se harán explícitos los criterios de calificación.

También se hará referencia en la programación de principio de curso a la forma de ofrecer las calificaciones al alumno. A cada alumno se le puede dar su calificación numérica, acompañada de una descripción cualitativa de la nota (aprobado, suspenso, etc.) y de recomendaciones escritas del profesor de cara a la corrección de errores, a posibles actividades de profundización...

En la fase de realización de la evaluación, se califica cada parte de la prueba por separado, atendiendo a los criterios y pesos relativos de cada uno. Así, la nota será igual a la suma de las puntuaciones conseguidas en cada parte.

3. Orientaciones sobre la escala a utilizar en la calificación y ponderación de los distintos elementos evaluativos.

Los exámenes tienen carácter liberatorio. Los alumnos recuperarán los objetivos-contenidos de las evaluaciones no superadas, tal y como aparece en el apartado siguiente. (recuperación).

Las pruebas de evaluación en este modelo son exclusivamente los exámenes.

La escala numérica comprenderá los valores de 0 a 10. El 5 es el valor mínimo necesario para aprobar.

4. Orientación para las actividades de recuperación.

La recuperación en el modelo B puede ser de dos formas y así constará en el programa:

a) Atendiendo al carácter continuo de la evaluación y sólo en el caso en que los objetivos-contenidos de una evaluación son requisitos previos para la superación de la siguiente, la nota en el primer examen será la que corresponda para cada alumno. La nota del segundo examen será la media aritmética de las notas conseguidas en el primer y segundo ejercicio. De esta manera, no se puede tener aprobada una evaluación sin tener superadas las anteriores.

b) En el caso en que no se dé la secuencia lógica en objetivos-contenidos que requiere este sistema de evaluación-recuperación, así como en el caso de que no se hayan recuperado las evaluaciones por el sistema anterior (porque estén todas suspensas, o porque habiendo aprobado las anteriores quede la última o últimas suspensas), la evaluación o evaluaciones no superadas se podrán recuperar:

b.1) en un examen final que tendrá lugar en junio.

b.2) en exámenes parciales durante el curso.

5. Tipo de información que se da al alumno antes de la evaluación.

En la programación se hará mención expresa de la información que se ha de dar a cada alumno o al grupo de clase, al principio de curso y antes de cada evaluación.

Se informará a los alumnos sobre los distintos aspectos de las pruebas de evaluación de forma detallada al principio de curso y antes de cada evaluación. Tal información se referirá a las características esenciales del modelo en cuestión y será distinta, en lo que a tipo de prueba se refiere, para el submodelo B1 y para el B2. Se informará al alumno sobre el número de evaluaciones, el modo de recuperación, la temporalización, los objetivos-contenidos asignados a cada una de las evaluaciones, el procedimiento general de evaluación, la forma de calificación, etc.

6. Uso de los datos de la evaluación.

El feed-back que se establece a partir de los resultados de la evaluación ha de ser lo más rápido posible (inmediato) y se hará de forma grupal, aunque, como se ha indicado anteriormente, la calificación numérica puede ir acompañada de una explicación y de recomendaciones breves, solamente más extensas en el caso de alumnos extremos.

Modelo C

Se identifica con el que habitualmente utilizan los profesores y se ha comprobado con la aplicación de un cuestionario previo. Tiene un carácter esencialmente sumativo y global. En la evaluación se tendrán en cuenta los objetivos de aprendizaje y las características de los estudiantes.

Para decidir sobre el grado de dominio alcanzado por los alumnos en una determinada materia, el profesor se puede basar indistintamente en el nivel de cada alumno en la asignatura en general, el trabajo de cada alumno en clase, el nivel de ejecución del alumno en cada una de las pruebas, etc.

A los alumnos se les ofrecerá una información previa adecuada acerca de la materia objeto de examen, temas o unidades a estudiar, nota mínima aproximada para alcanzar el aprobado, etc.

Entrenamiento

El entrenamiento consistió, básicamente, en un curso intensivo, impartido a los profesores seleccionados para llevar a cabo la experiencia. El curso se estructuró en módulos comunes para todos los profesores y en módulos específicos para cada modelo, excepto para el modelo C que por definición era grupo de control. Antes de proceder al entrenamiento fue aplicado un cuestionario sobre las características de la evaluación en 8º de EGB tal y como eran percibidas por los profesores. La información proporcionada por este cuestionario sirvió de base para constatar las diferencias entre las formas de evaluación utilizadas por el profesor y las exigidas por los modelos.

La aplicación de los modelos evaluativos fue objeto de seguimiento por el equipo investigador con la finalidad de constatar el grado en que se cumplía el programa establecido.

RESULTADOS

Presentamos, en primer lugar, una descripción general de los datos obtenidos, sintetizados en medidas de tendencia central y de variabilidad. En una segunda fase se ofrecen los resultados del análisis primario que se concreta en la determinación de significación de diferencias en el rendimiento lingüístico y matemático entre los distintos modelos evaluativos, utilizando para ello el análisis de covarianza.

1) Descripción general

En las tablas que siguen se presentan las medias y desviaciones típicas en todas las variables medidas, por modelos evaluativos (tabla 1), tipo de centro (tabla 2), y sexo (tabla 3).

Tabla 1
MEDIAS Y DESVIACIONES TÍPICAS EN LAS DISTINTAS VARIABLES
MEDIDAS, OBTENIDAS POR LOS GRUPOS QUE SIGUIERON LOS
DISTINTOS MODELOS EVALUATIVOS

		Variables								
Modelos		PAL	SIM	ANA	CL	GEFT	PLE	PMAT	ALE	AMA
A	\bar{x}	4.16	6.80	9.59	14.85	10.42	20.04	12.99	53.55	51.83
	s	1.57	2.70	2.99	3.72	4.70	3.72	3.79	11.89	9.34
B1	\bar{x}	2.38	7.53	11.34	14.50	11.11	19.51	15.72	53.43	54.48
	s	1.01	3.60	2.72	3.84	5.11	3.55	4.14	11.78	9.34
B2	\bar{x}	2.63	6.75	9.19	13.26	9.64	19.60	13.51	56.27	52.66
	s	1.24	2.42	2.58	3.40	4.81	4.21	4.15	11.46	10.40
C	\bar{x}	2.63	7.19	9.01	12.87	9.87	16.26	12.11	53.79	53.16
	s	1.47	3.62	3.19	3.87	7.04	4.78	4.88	11.20	10.06
	\bar{x}_g	2.97	7.04	9.64	13.78	10.23	19.17	13.54	54.16	52.95
	Sg	1.53	3.12	3.02	3.79	5.57	4.21	4.46	11.62	9.80

Variables

- PAL. Palabras
- SIM. Símbolos
- ANA. Analogías
- CL. Comprensión lectora
- GEFT. DIC
- PLE. Prueba de Lengua
- PMAT. Prueba de Matemáticas
- ALE. Actitud Lengua
- AMA. Actitud Matemáticas

En la variable PAL la media más alta corresponde al modelo A y la más baja al modelo B1. Por encima de la media sólo se encuentra el modelo A. La variabilidad más alta corresponde al modelo A y la única por encima de la media.

La variable SIM obtiene la media más alta en el modelo B1, y por encima de la media resultan las puntuaciones obtenidas por el modelo B1 y C. La más alta variabilidad corresponde al C y por encima de la general está también el B1.

En la variable ANA, la \bar{x} más alta y la única por encima de la media corresponde

al modelo B1. La variabilidad más alta es la del modelo C y la única por encima de la general.

En CL, la media más alta corresponde al modelo A, quedando por encima de la media también el B1, y la variabilidad más alta es la del C y por encima de la media también la presenta el B1.

En la variable GEFT, la máxima puntuación es obtenida por el modelo B1, estando por encima de la media general este modelo y el A. La máxima variabilidad la obtiene el modelo C y el B1.

En la variable PLE, vuelve a obtener la más alta puntuación el modelo A, seguido por el B2 y B1, todos por encima de la media general. La variabilidad más alta corresponde al modelo C.

En la prueba de Matemáticas PMAT, la media más alta es la obtenida por el el modelo B1, y la mayor variabilidad es la del modelo C.

En ALE la media más alta corresponde al modelo B2, y la máxima variabilidad, al modelo A.

En AMA, la media más alta, corresponde a B1, y la mayor variabilidad al B2.

En resumen, el modelo B1 alcanza la media más alta en cinco variables, el modelo A en otras tres (relacionadas con la lengua) y el modelo B2 en una.

El modelo C tiene la mayor variabilidad en 6 variables, el modelo A en dos (relacionadas con la Lengua) y el B2 en una.

Tabla 2
MEDIDAS Y DESVIACIONES TÍPICAS POR TIPO DE CENTRO

Variables

Tipo de centro		PAL	SIM	ANA	CL	GEFT	PLE	PMAT	ALE	AMA
Público	\bar{x}	2.79	6.89	9.09	13.47	9.83	18.32	12.92	53.84	52.69
	s	1.44	3.20	2.94	3.78	6.27	4.35	4.26	11.74	10.00
Privado	\bar{x}	3.11	7.16	10.13	14.08	10.54	19.63	13.92	54.39	53.14
	s	1.59	3.05	3.01	3.78	4.93	4.06	4.53	11.53	9.66
	\bar{x}_g	2.97	7.04	9.64	13.78	10.23	19.17	13.54	54.16	52.95
	Sg	1.53	3.12	3.02	3.79	5.57	4.21	4.46	11.62	9.80

La media es más alta en los centros privados para todas las variables. La mayor dispersión corresponde a los colegios públicos en cinco de las nueve variables.

Tabla 3
MEDIAS Y DESVIACIONES TÍPICAS POR SEXO

Variables

Sexo		PAL	SIM	ANA	CL	GEFT	PLE	PMAT	ALE	AMA
Hombres	\bar{x}	2.71	7.25	9.93	13.31	10.20	18.22	13.43	52.50	53.02
	s	1.51	3.45	3.14	3.81	4.98	4.29	4.57	11.15	9.45
Mujeres	\bar{x}	3.25	6.82	9.34	14.21	10.28	20.32	13.71	56.25	53.07
	s	1.51	2.74	2.87	3.73	6.18	3.74	4.31	11.62	9.91
	\bar{x}_g	2.97	7.04	9.64	13.78	10.23	19.17	13.54	54.16	52.95
	Sg	1.53	3.12	3.02	3.79	5.57	4.21	4.46	11.62	9.80

El que la \bar{x}_g sea más baja que la del grupo de mujeres y hombres obedece al hecho del que hay una pequeña diferencia en la muestra de cada grupo (hombres y mujeres) y el número total de individuos. Esta diferencia puede deberse al no haber podido adjudicar a uno de los dos sexos algunas de las pruebas realizadas. (Ver puntuaciones de la variable AMA).

Las medias de las mujeres son más altas en 7 variables. La de los hombres es más alta en SIM y en ANA. La mayor dispersión corresponde a los hombres en cinco variables y a las mujeres en tres.

II. DIFERENCIAS EN RENDIMIENTOS EN LENGUA Y MATEMÁTICAS ENTRE LOS DISTINTOS MODELOS EVALUATIVOS

Dadas las características de los datos limitados a los resultados de aplicar los instrumentos de medida del rendimiento en Lengua y Matemáticas al final del tratamiento y de variables aptitudinales presumiblemente relacionadas con el mismo durante el tratamiento, parecía aconsejable proceder al contraste de la hipótesis general, es decir, la existencia de diferencias significativas en el rendimiento en Lengua y Matemáticas entre los grupos que siguieron modelos evaluativos distintos. Para ello, nos pareció que la prueba estadística adecuada sería el análisis de covarianza introduciendo como covariables sucesivamente una, dos o más de las variables representativas de las aptitudes medidas.

A. Diferencia en rendimiento en Lengua

Se presentan los resultados del análisis de covarianza, usando como covariables:

PAL (Tablas 4, 5 y 6)

PAL y CL (Tablas 7, 8 y 9)

PAL, SIM y CL (Tablas 10, 11 y 12)

PAL, SIM, ANA y CL (Tablas 13, 14 y 15)

PAL, SIM, ANA, CL y GEFT (Tablas 16, 17 y 18)

Diferencias en los Modelos en rendimiento en Lengua usando como covariable «PAL».

Tabla 4
MEDIAS Y MEDIAS AJUSTADAS

	A	B1	B2	C
\bar{x} PAL	4.35	2.81	2.52	2.32
\bar{x} PLE	20.07	20.63	17.95	16.37
\bar{x} ajustadas PLE	19.48	20.79	18.25	16.77

Tabla 5
TABLA DE ANÁLISIS DE COVARIANZA

<i>Fuentes de variación</i>	S.C.	g.l.	M.C.	F	sig.
<i>Entre modelos</i>	979.94	3	326.65	22.25	0.000
<i>Error</i>	7737.39	527	14.68		

Tabla 6
MATRIZ DE PRUEBA T ENTRE LAS MEDIAS AJUSTADAS DE LOS
DISTINTOS MODELOS

	A	B1	B2	C
A	0.00 ($\alpha=1.000$)			
B1	2.53 ($\alpha=0.012$)	0.00 ($\alpha=1.000$)		
B2	2.40 ($\alpha=0.016$)	4.86 ($\alpha=0.000$)	0.00 ($\alpha=1.000$)	
C	5.34 ($\alpha=0.000$)	7.90 ($\alpha=0.000$)	3.06 ($\alpha=0.002$)	0.00 ($\alpha=1.000$)

Introduciendo la covariable PAL, aparecen diferencias altamente significativas en rendimiento entre los distintos modelos (Tabla 5). En consecuencia, en este caso, se rechaza la hipótesis nula y se confirma la hipótesis general, es decir, hay diferencias significativas en rendimiento entre los distintos modelos.

Realizados contrastes posteriores (Tabla 6) se observa que las diferencias más significativas son las existentes entre los modelos B1 y C, y después entre los A y C, como ya aparecía en la fila de medias ajustadas de la tabla 4.

Diferencias en los Modelos en rendimiento en Lengua usando como covariable «PAL» y «CL»

Tabla 7
MEDIAS Y MEDIAS AJUSTADAS

	A	B1	B2	C
\bar{x} PAL	4.35	2.81	2.52	2.32
\bar{x} CL	15.29	14.54	12.63	12.26
\bar{x} PLE	20.07	20.63	17.95	16.37
\bar{x} ajustadas PLE	19.07	20.45	18.60	17.24

Tabla 8
TABLA DE ANÁLISIS DE COVARIANZA

<i>Fuentes de variación</i>	S.C.	g.l.	M.C.	F	sig
<i>Entre modelos</i>	569.38	3	189.79	14.82	0.000
<i>Error</i>	6737.37	526	12.81		

Tabla 9
MATRIZ DE PRUEBAS T ENTRE LAS MEDIAS AJUSTADAS DE LOS
DISTINTOS MODELOS

	A	B1	B2	C
A	0.00 ($\&$ =1.000)			
B1	2.85 ($\&$ =0.004)	0.00 ($\&$ =1.000)		
B2	0.97 ($\&$ =0.331)	3.75 ($\&$ =0.000)	0.00 ($\&$ =1.000)	
C	3.79 ($\&$ =0.000)	6.64 ($\&$ =0.000)	3.02 ($\&$ =0.002)	0.00 ($\&$ =1.000)

Introduciendo las covariables PAL y CL, aparecen diferencias altamente significativas entre los distintos modelos (tabla 8). En consecuencia, en este caso, se rechaza la hipótesis nula y se confirma la hipótesis general, es decir, hay diferencias significativas en rendimiento entre los distintos modelos.

Realizando contrastes posteriores (tabla 9) se observa que las diferencias más significativas son las existentes entre los modelos B1 y C y entre el A y el C, como ya aparecía en la línea de medias ajustadas de la tabla 7.

Diferencias en los Modelos en rendimiento en Lengua usando como covariables «PAL», «SIM» y «CL»

Tabla 10
MEDIAS Y MEDIAS AJUSTADAS

	A	B1	B2	C
\bar{x} PAL	4.35	2.81	2.52	2.32
\bar{x} SIM	7.04	9.02	5.99	6.63
\bar{x} CL	15.29	14.54	12.63	12.26
\bar{x} PLE	20.07	20.63	17.95	16.37
\bar{x} ajustadas PLE	19.23	19.97	18.79	17.23

Tabla 11
TABLA DE ANÁLISIS DE COVARIANZA

<i>Fuentes de variación</i>	S.C.	g.l.	M.C.	F	sig
<i>Entre modelos</i>	437.41	3	145.80	11.82	0.000
<i>Error</i>	6476.93	525	12.34		

Tabla 12
MATRIZ DE PRUEBAS T ENTRE LAS MEDIAS AJUSTADAS DE LOS
DISTINTOS MODELOS

	A	B1	B2	C
A	0.00 ($\&=1.000$)			
B1	1.51 ($\&=0.132$)	0.00 ($\&=1.000$)		
B2	0.93 ($\&=0.353$)	2.35 ($\&=0.019$)	0.00 ($\&=1.000$)	
C	4.20 ($\&=0.000$)	5.65 ($\&=0.000$)	3.50 ($\&=0.000$)	0.00 ($\&=1.000$)

Introduciendo la covariable SIM al análisis anterior, aparecen también diferencias altamente significativas en rendimiento entre los distintos modelos (tabla 11). En consecuencia, en este caso, se rechaza la hipótesis nula y se confirma la hipótesis general, es decir, hay diferencias significativas en rendimiento entre los distintos modelos.

Realizados contrastes posteriores (tabla 12), se observa que las diferencias más significativas aparecen también entre los modelos B1 y C, y entre A y C, como ya aparece en la línea de medias ajustadas de la tabla 10.

Diferencias en los Modelos en rendimiento en Lengua usando como covariables «PAL», «SIM», «ANA» y «CL»

Tabla 13
MEDIAS Y MEDIAS AJUSTADAS

	A	B1	B2	C
\bar{x} PAL	4.35	2.81	2.52	2.32
\bar{x} SIM	7.04	9.02	5.99	6.63
\bar{x} ANA	9.59	11.93	8.49	8.17
\bar{x} CL	15.29	14.54	12.63	12.26
\bar{x} PLE	20.07	20.63	17.95	16.37
\bar{x} ajustadas PLE	19.33	19.43	18.89	17.40

Tabla 14
TABLA DE ANÁLISIS DE COVARIANZA

<i>Fuentes de variación</i>	S.C.	g.l.	M.C.	F	sig
<i>Entre modelos</i>	284.22	3	94.74	7.90	0.000
<i>Error</i>	6281.17	524	11.99		

Tabla 15
MATRIZ DE PRUEBAS T ENTRE LAS MEDIAS AJUSTADAS DE LOS
DISTINTOS MODELOS

	A	B1	B2	C
A	0.00 ($\alpha=1.000$)			
B1	0.21 ($\alpha=0.836$)	0.00 ($\alpha=1.000$)		
B2	0.93 ($\alpha=0.353$)	1.03 ($\alpha=0.300$)	0.00 ($\alpha=1.000$)	
C	4.10 ($\alpha=0.000$)	3.98 ($\alpha=0.000$)	3.39 ($\alpha=0.000$)	0.00 ($\alpha=1.000$)

Al introducir una covariable más, ANA, aparecen diferencias altamente significativas en el rendimiento de los distintos modelos (tabla 14). En consecuencia, en este caso, se rechaza la hipótesis nula y se confirma la hipótesis general, es decir, hay diferencias significativas entre los distintos modelos.

Realizados contrastes posteriores (tabla 15) se observa que las diferencias más significativas se dan entre A y C y entre B1 y C, como aparece en la línea de medias ajustadas de la tabla 13.

Diferencias en los Modelos en rendimiento en Lengua usando como covariables «PAL», «SIM», «ANA», «CL» y «GEFT»

Tabla 16
MEDIAS Y MEDIAS AJUSTADAS

	A	B1	B2	C
\bar{x} PAL	4.35	2.81	2.52	2.32
\bar{x} SIM	7.04	9.02	5.99	6.63
\bar{x} ANA	9.59	11.93	8.49	8.17
\bar{x} CL	15.29	14.54	12.63	12.26
\bar{x} GEFT	10.76	13.00	8.58	10.28
\bar{x} PLE	20.07	20.63	17.95	16.37
\bar{x} ajustadas PLE	19.32	19.37	18.98	17.38

Tabla 17
TABLA DE ANÁLISIS DE COVARIANZA

<i>Fuentes de variación</i>	S.C.	g.l.	M.C.	F	sig
<i>Entre modelos</i>	290.35	3	96.78	8.16	0.000
<i>Error</i>	6205.49	523	11.86		

Tabla 18
MATRIZ DE PRUEBAS T ENTRE LAS MEDIAS AJUSTADAS DE LOS
DISTINTOS MODELOS

	A	B1	B2	C
A	0.00 (&=1.000)			
B1	0.09 (&=0.928)	0.00 (&=1.000)		
B2	0.73 (&=0.463)	0.74 (&=0.458)	0.00 (&=1.000)	
C	4.15 (&=0.000)	3.91 (&=0.000)	3.64 (&=0.000)	0.00 (&=1.000)

Con la covariable GEFT, también aparecen diferencias altamente significativas en el rendimiento de los distintos modelos (Tabla 17). En consecuencia, en este caso, se rechaza la hipótesis nula y se confirma la hipótesis general, es decir, hay diferencias significativas entre los distintos modelos.

Realizados contrastes posteriores (Tabla 18) se observa que las diferencias más significativas se dan entre A y C y entre B1 y C como aparece en la línea de medias ajustadas de la tabla 16.

En resumen, las diferencias más significativas en rendimiento en Lengua se encuentran entre los modelos B1 y C en tres análisis con las covariables PAL, PAL y CL, y PAL, CL y SIM.

Al introducir las otras dos covariables, ANA y GEFT, el nivel de significatividad es superior entre A y C, aunque las diferencias son ligeramente mayores entre B1 y C.

B) Diferencias en rendimiento en Matemáticas

Se presentan los resultados del análisis de covarianza usando como covariables:
PAL (Tablas 19, 20 y 21)
SIM (Tablas 22, 23 y 24)
ANA (Tablas 25, 26 y 27)

SIM y ANA (Tablas 28, 29 y 30)

SIM, ANA y GEFT (Tablas 31, 32 y 33)

PAL, SIM, ANA, CL y GEFT (Tablas 34, 35 y 36)

Diferencias en los Modelos en rendimiento en Matemáticas usando como covariable «PAL»

Tabla 19
MEDIAS Y MEDIAS AJUSTADAS

	A	B1	B2	C
\bar{x} PAL	4.35	2.81	2.52	2.32
\bar{x} PMAT	13.46	17.95	12.05	9.98
\bar{x} ajustadas PMAT	13.02	18.07	12.27	10.28

Tabla 20
TABLA DE ANÁLISIS DE COVARIANZA

<i>Fuentes de variación</i>	S.C.	g.l.	M.C.	F	sig
<i>Entre modelos</i>	3600.83	3	1200.28	93.36	0.000
<i>Error</i>	6775.22	527	2.86		

Tabla 21
MATRIZ DE PRUEBAS T ENTRE LAS MEDIAS AJUSTADAS DE LOS
DISTINTOS MODELOS

	A	B1	B2	C
A	0.00 ($\alpha=1.000$)			
B1	10.42 ($\alpha=0.000$)	0.00 ($\alpha=1.000$)		
B2	1.56 ($\alpha=0.119$)	11.85 ($\alpha=0.000$)	0.00 ($\alpha=1.000$)	
C	5.78 ($\alpha=0.000$)	16.36 ($\alpha=0.000$)	4.41 ($\alpha=0.000$)	0.00 ($\alpha=1.000$)

Introduciendo la covariable PAL, aparecen diferencias altamente significativas en rendimiento en los distintos modelos (Tabla 20). En consecuencia, en este caso, se rechaza la hipótesis nula y se confirma la hipótesis general, es decir, hay diferencias significativas en rendimiento entre los distintos modelos.

Realizando contrastes posteriores (Tabla 21), se observa que las diferencias más significativas son las existentes entre los modelos B1 y C y después entre B1 y B2, como ya aparece en las medias ajustadas de la tabla 19.

Diferencias en los Modelos en rendimiento en Matemáticas usando como covariable «SIM»

Tabla 22
MEDIAS Y MEDIAS AJUSTADAS

	A	B1	B2	C
\bar{x} SIM	7.04	9.02	5.99	6.63
\bar{x} PMAT	13.46	17.95	12.05	9.98
\bar{x} ajustadas PMAT	13.47	17.14	12.50	10.17

Tabla 23
TABLA DE ANÁLISIS DE COVARIANZA

<i>Fuentes de variación</i>	S.C.	g.l.	M.C.	F	sig
<i>Entre modelos</i>	2646.56	3	882.19	75.75	0.000
<i>Error</i>	6136.98	527	11.64		

Tabla 24
MATRIZ DE PRUEBAS T ENTRE LAS MEDIAS AJUSTADAS DE LOS
DISTINTOS MODELOS

	A	B1	B2	C
A	0.00 ($\&=1.000$)			
B1	8.36 ($\&=0.000$)	0.00 ($\&=1.000$)		
B2	2.37 ($\&=0.179$)	9.47 ($\&=0.000$)	0.00 ($\&=1.000$)	
C	8.49 ($\&=0.000$)	14.95 ($\&=0.000$)	5.42 ($\&=0.000$)	0.00 ($\&=1.000$)

Con la covariable SIM aparecen diferencias altamente significativas en rendimiento en los distintos modelos (Tabla 23). En consecuencia, se rechaza la hipótesis nula y se confirma la hipótesis general, es decir, hay diferencias significativas en rendimiento entre los distintos modelos.

Realizando contrastes posteriores (Tabla 24), se observa que las diferencias más significativas son las existentes entre los modelos B1 y C y después entre B1 y B2, como ya aparece en las medias ajustadas de la tabla 22.

Diferencias en los Modelos en rendimiento en Matemáticas usando como covariable «ANA»

Tabla 25
MEDIAS Y MEDIAS AJUSTADAS

	A	B1	B2	C
\bar{x} ANA	9.59	11.93	8.49	8.17
\bar{x} PMAT	13.46	17.95	12.05	9.98
\bar{x} ajustadas PMAT	13.40	17.02	12.40	10.45

Tabla 26
TABLA DE ANÁLISIS DE COVARIANZA

<i>Fuentes de variación</i>	S.C.	g.l.	M.C.	F	sig
<i>Entre modelos</i>	2079.59	3	693.19	57.42	0.000
<i>Error</i>	6362.39	527	12.07		

Tabla 27
MATRIZ DE PRUEBAS T ENTRE LAS MEDIAS AJUSTADAS DE LOS DISTINTOS MODELOS

	A	B1	B2	C
A	0.00 ($\&=1.000$)			
B1	7.98 ($\&=0.000$)	0.00 ($\&=1.000$)		
B2	2.40 ($\&=0.017$)	9.05 ($\&=0.000$)	0.00 ($\&=1.000$)	
C	7.30 ($\&=0.000$)	13.03 ($\&=0.000$)	4.45 ($\&=0.000$)	0.00 ($\&=1.000$)

Usando como covariable ANA, aparecen también diferencias altamente significativas en rendimiento en los distintos modelos (Tabla 26). En consecuencia, en este caso, se rechaza la hipótesis nula y se confirma la hipótesis general, es decir, hay diferencias significativas en rendimiento entre los distintos modelos.

Realizando contrastes posteriores (Tabla 27), se observa que las diferencias más significativas son las existentes entre los modelos B1 y C y después entre B1 y B2, como ya aparece en las medias ajustadas de la tabla 25.

Diferencias en los Modelos en rendimiento en Matemáticas usando como covariables «SIM» y «ANA»

Tabla 28
MEDIAS Y MEDIAS AJUSTADAS

	A	B1	B2	C
\bar{x} SIM	7.04	9.02	5.99	6.63
\bar{x} ANA	9.59	11.93	8.49	8.17
\bar{x} PMAT	13.46	17.95	12.05	9.98
\bar{x} ajustadas PMAT	13.42	16.50	12.72	10.52

Tabla 29
TABLA DE ANÁLISIS DE COVARIANZA

<i>Fuentes de variación</i>	S.C.	g.l.	M.C.	F	sig
<i>Entre modelos</i>	1677.94	3	559.31	50.61	0.000
<i>Error</i>	5812.87	526	11.05		

Tabla 30
MATRIZ DE PRUEBAS T ENTRE LAS MEDIAS AJUSTADAS DE LOS
DISTINTOS MODELOS

	A	B1	B2	C
A	0.00 ($\&=1.000$)			
B1	6.99 ($\&=0.000$)	0.00 ($\&=1.000$)		
B2	1.75 ($\&=0.080$)	7.53 ($\&=0.000$)	0.00 ($\&=1.000$)	
C	7.52 ($\&=0.000$)	12.24 ($\&=0.000$)	5.24 ($\&=0.000$)	0.00 ($\&=1.000$)

Introduciendo dos covariables (SIM y ANA) aparecen también diferencias altamente significativas en rendimiento en los distintos modelos (Tabla 29). En consecuencia, en este caso, se rechaza la hipótesis nula y se confirma la hipótesis general, es decir, hay diferencias significativas en rendimiento entre los distintos modelos.

Realizando contrastes posteriores (Tabla 30), se observa que las diferencias más significativas son las existentes entre los modelos B1 y C y después entre B1 y B2, como aparece en las medias ajustadas de la tabla 28.

Diferencias en los Modelos en rendimiento en Matemáticas usando como covariables «SIM», «ANA» y GEFT»

Tabla 31
MEDIAS Y MEDIAS AJUSTADAS

	A	B1	B2	C
\bar{x} SIM	7.04	9.02	5.99	6.63
\bar{x} ANA	9.59	11.93	8.49	8.17
\bar{x} WIT	10.76	13.00	8.58	10.28
\bar{x} PMAT	13.46	17.95	12.05	9.98
\bar{x} ajustadas PMAT	13.41	16.44	12.80	10.51

Tabla 32
TABLA DE ANÁLISIS DE COVARIANZA

<i>Fuentes de variación</i>	S.C.	g.l.	M.C.	F	sig
<i>Entre modelos</i>	1651.25	3	550.42	50.23	0.000
<i>Error</i>	5752.98	525	10.96		

Tabla 33
MATRIZ DE PRUEBAS T ENTRE LAS MEDIAS AJUSTADAS DE LOS
DISTINTOS MODELOS

	A	B1	B2	C
A	0.00 ($\alpha=1.000$)			
B1	6.89 ($\alpha=0.000$)	0.00 ($\alpha=1.000$)		
B2	1.53 ($\alpha=0.125$)	7.23 ($\alpha=0.000$)	0.00 ($\alpha=1.000$)	
C	7.56 ($\alpha=0.000$)	12.18 ($\alpha=0.000$)	5.46 ($\alpha=0.000$)	0.00 ($\alpha=1.000$)

Añadiendo la covariable GEFT también aparecen diferencias altamente significativas en rendimiento en los distintos modelos (Tabla 32). En consecuencia, en este caso, se rechaza la hipótesis nula y se confirma la hipótesis general, es decir, hay diferencias significativas en rendimiento entre los distintos modelos.

Realizando contrastes posteriores (Tabla 32) se observa que las diferencias más significativas son las existentes entre los modelos B1 y C y después entre A y C, como aparece en las medias ajustadas de la tabla 31.

Diferencias en los Modelos en rendimiento en Matemáticas usando como covariables «PAL», «SIM», «ANA» «CL» y «GEFT»

Tabla 34
MEDIAS Y MEDIAS AJUSTADAS

	A	B1	B2	C
\bar{x} PAL	4.35	2.81	2.52	2.32
\bar{x} SIM	7.04	9.02	5.99	6.63
\bar{x} ANA	9.59	11.93	8.49	8.17
\bar{x} CL	15.29	14.54	12.63	12.26
\bar{x} WIT	10.76	13.00	8.58	10.28
\bar{x} PMAT	13.46	17.95	12.05	9.98
\bar{x} ajustadas PMAT	12.98	16.62	12.95	10.81

Tabla 35
TABLA DE ANÁLISIS DE COVARIANZA

<i>Fuentes de variación</i>	S.C.	g.l.	M.C.	F	sig
<i>Entre modelos</i>	1566.66	3	522.22	49.99	0.000
<i>Error</i>	5463.13	523	10.45		

Tabla 36
**MATRIZ DE PRUEBAS T ENTRE LAS MEDIDAS AJUSTADAS DE LOS
 DISTINTOS MODELOS**

	A	B1	B2	C
A	0.00 ($\alpha=1.000$)			
B1	7.61 ($\alpha=0.000$)	0.00 ($\alpha=1.000$)		
B2	0.06 ($\alpha=0.948$)	7.46 ($\alpha=0.000$)	0.00 ($\alpha=1.000$)	
C	4.96 ($\alpha=0.000$)	12.21 ($\alpha=0.000$)	5.22 ($\alpha=0.000$)	0.00 ($\alpha=1.000$)

Al introducir las covariables CL y PAL las diferencias en rendimiento entre los distintos modelos son altamente significativas (Tabla 35). En consecuencia, en este caso, se rechaza la hipótesis nula y se confirma la hipótesis general, es decir, hay diferencias significativas en rendimiento entre los distintos modelos.

Realizando contrastes posteriores (Tabla 36), se observa que las diferencias más significativas son las existentes entre los modelos B1 y C y después entre los A y B1 y B1 y B2, como se observa también en las medias ajustadas de la tabla 34.

En síntesis, las diferencias más significativas en rendimiento en matemáticas se encuentran entre los modelos B1 y C con cualquier covariable introducida. En segundo lugar, las diferencias mayores se dan entre B1 y B2 en tres, cuatro casos y entre A y C en uno.

Para ambas variables (rendimiento en Lengua y rendimiento en Matemáticas) las diferencias más significativas se dan entre los modelos B1 y C.

En resumen, como muestran las tablas 37 y 38 los resultados pueden sintetizarse así:

Tabla 37
 DIFERENCIAS DE RENDIMIENTO EN LENGUA ENTRE GRUPOS CON
 MODELOS EVALUATIVOS DISTINTOS AJUSTANDO LAS MEDIAS EN
 FUNCIÓN DE DIVERSAS COVARIABLES

Diferencias entre modelos

Covariables	General			dos a dos		
	A-B1	A-B2	A-C	B1-B2	B1-C	B2-C
	PAL	***	*	*	***	***
PAL, CL	***	**	—	***	***	***
PAL, SIM, CL	***	—	—	***	*	***
PAL, SIM, ANA, CL	***	—	—	***	—	***
PAL, SIM, SANA, CL, GEFT	***	—	—	***	—	***

Nivel de significación:

- * $p \leq .05$
- ** $p \leq .01$
- *** $p \leq .001$

Tabla 38
**DIFERENCIAS DE RENDIMIENTO EN MATEMÁTICAS ENTRE GRUPOS
 CON MODELOS EVALUATIVOS DISTINTOS AJUSTANDO LAS MEDIAS
 EN FUNCIÓN DE DIVERSAS COVARIABLES**

Diferencias entre modelos

Covariables	General			dos a dos		
	A-B1	A-B2	A-C	B1-B2	B1-C	B2-C
PAL	***	***	—	***	***	***
SIM	***	***	—	***	***	***
ANA	***	***	*	***	***	***
SIM, ANA	***	***	—	***	***	***
SIM, ANA, GEFT	***	***	—	***	***	***
PAL, SIM, SANA, CL, GEFT	***	***	—	***	***	***

Nivel de significación:

- * $p \leq .05$
- ** $p \leq .01$
- *** $p \leq .001$

La evaluación afecta al rendimiento en las materias de Lengua y Matemáticas cuando se consideran los modelos experimentados en el presente estudio:

Modelo A: Evaluación continua y formativa.

Modelo B: Evaluación criterial y formativa.

Submodelo B1.- Prueba objetiva

Submodelo B2.- Preguntas cortas

Modelo C: Evaluación sumativa. Modelo control (Ev. normalmente realizada por el profesor)

Sin embargo, la mayor parte de las diferencias se deben a la superioridad de los modelos B1, B2 y A sobre el C. En Lengua aparecen diferencias en rendimiento cuando se comparan los efectos de A con B2, muy escasos entre A y B1 y sensiblemente marcados, aunque no generales, entre B1 y B2 aunque en este último caso la

diferencia sólo es altamente significativa cuando se elimina el efecto en el rendimiento de las variables PAL y PAL y CL.

El modelo evaluativo vinculado a rendimientos más altos en Lengua es el B1, después sigue el A y el B2 con una pequeña superioridad a favor del modelo A, aunque como hemos visto no hay diferencias significativas entre A y B2.

Resulta obvio que el modelo evaluativo con menores efectos optimizantes en el rendimiento de Lengua es el C, que hemos utilizado como modelo control en el presente estudio.

En Matemáticas no existen diferencias entre los efectos en rendimiento producidos por los modelos A y B2, excepto en el caso en que se elimina el efecto debido a la covariable ANA en que las diferencias a favor de A son significativas a nivel de probabilidad .05

El modelo evaluativo vinculado a rendimiento más alto en Matemáticas es también el B1, después sigue el A y el B2 con una pequeña superioridad a favor del modelo A.

Igual que en Lengua el modelo evaluativo con menor efecto optimizante en el rendimiento en Matemáticas es el modelo C utilizado como control. Quizá haya que pensar en el efecto «Hawthorne» como explicación, en parte, de los resultados en los grupos que aplicaron los modelos propiamente experimentales (A, B1 y B2).

CONCLUSIONES

Pese a la gran dificultad de control en experimentos de campo realizados en contextos sociales tan complejos como los centros educativos, se ha confirmado, sin lugar a dudas, la hipótesis básica del estudio que hace referencia al impacto de la evaluación en el rendimiento educativo. Este hecho reafirma la gran intensidad del influjo evaluativo en el producto de la educación.

Aunque las condiciones de realización del experimento y otros factores no contemplados en el diseño, impidieron la aplicación de pretest y otras pruebas programadas, los datos finalmente obtenidos permitieron un análisis de covarianza que ha puesto de manifiesto diferencias altamente significativas en el rendimiento de Lengua y Matemáticas, asociadas a la utilización de distintos modelos evaluativos.

Estos resultados no parece que puedan explicarse por los conocimientos previos de los alumnos, efecto a su vez de la variabilidad de rendimiento entre los distintos centros educativos, ya que todos los modelos fueron probados en todos los tipos de centros: privados y públicos y de diferentes contextos socio-económicos.

Tampoco tenemos evidencia de que los resultados pudieran estar estrechamente asociados a la capacidad y tipo de profesor ya que cada uno de los modelos evaluativos fue asignado a un número de profesores (entre 8 y 10) que presumiblemente asegura un nivel suficiente de representatividad del grupo.

Con estas premisas, las conclusiones que ofrecemos a continuación parecen apoyadas en una base sólida de evidencia empírica:

- 1) Parece que los modelos evaluativos con énfasis en la función formativa y en la referencia criterial resultan más eficaces que los orientados a la función sumativa y con referencia normativa; sin embargo, la clara distinción a lo largo del proceso evaluativo de los aspectos formativo y sumativo, así como el conocimiento previo de las consecuencias de los resultados de la evaluación aparecen como determinantes de su impacto en el rendimiento.
- 2) En cuanto a la frecuencia de la evaluación, resultan superiores los modelos con un número fijo de evaluaciones previamente establecido en la programación y con carácter liberatorio a aquéllos otros que ponen énfasis en la continuidad sin establecer momentos precisos de evaluación formal.
- 3) La superioridad señalada en el apartado anterior aparece asociada al tipo de pruebas utilizado en las sesiones de evaluación programadas. Concretamente, resulta significativamente superior el modelo que utiliza pruebas objetivas tanto en Matemáticas como en Lengua, complementadas por resolución de problemas en la primera y comentario de textos en la segunda.
- 4) No ha aparecido suficiente evidencia acerca del impacto del rendimiento en Lengua y Matemáticas de las diferencias de la evaluación en los siguientes aspectos:
 - a) Definición de criterios o indicadores de calificación de cada prueba.
 - b) Orientaciones sobre la escala a utilizar en la calificación y ponderación de los distintos elementos evaluativos.
 - c) Orientaciones para las actividades de recuperación.
 - d) Tipo de información que se proporciona al alumno antes de la prueba.
 - e) Uso de los datos de la evaluación.

LIMITACIONES DEL ESTUDIO

Las limitaciones en la investigación experimental en Educación derivan fundamentalmente de la dificultad de control. La complejidad de las situaciones educativas en contextos institucionales constituye un desafío para el diseño de experimentos con pretensiones de validez.

La aplicación de algunos niveles de nuestra variable experimental —modelos de evaluación— no es posible en algunos centros escolares sin modificar radicalmente los modos de hacer, los valores, los principios pedagógicos, las relaciones con las familias, etc., es decir, sin cambiar la cultura de la institución. En estas circunstancias, debemos aceptar que los resultados educativos —variable dependiente— sea efecto, en cierto grado, de las características inmodificables de cada escuela o de la interacción entre evaluación —tratamiento— y cultura institucional.

Por otra parte, la imposibilidad de aplicar un pretest para determinar los conocimientos previos en Lengua y Matemáticas nos ha impedido disponer en el análisis de una covariable que hubiera podido paliar el riesgo de invalidez derivado de la incapacidad del diseño apuntada en el párrafo anterior.

Finalmente, es necesario aludir también al hecho de no haber podido controlar totalmente el efecto del profesor en la variable dependiente, ya que el proceso de asignación de profesores a modelos no fue completamente aleatorio. Los profesores de cada institución hubieron de adoptar el modelo asignado al Centro.

ALGUNAS SUGERENCIAS

Dadas las dificultades de control inherentes a los experimentos en situaciones educativas reales, señaladas anteriormente, parece que una vía para superarlas sería la puesta en práctica de todos los tratamientos objeto de comparación —los diversos niveles de la variable experimental— en cada uno de los centros elegidos. Este diseño aseguraría el presupuesto básico de la experimentación científica: mantener iguales todas las circunstancias, excepto el tratamiento, es decir, homogeneizar el efecto de la multitud de variables intervinientes sobre la variable dependiente, para que puedan manifestarse nítidamente, sin enmascaramientos, los efectos diferenciales de los distintos niveles del tratamiento en la variable dependiente.

Naturalmente, esto no es posible cuando el tratamiento se identifica con un componente esencial del proceso de intervención educativa, como es el modelo de evaluación adoptado. Un modelo de evaluación supone en última instancia un modelo de instrucción e, incluso, de educación y, obviamente, en una misma institución no son compatibles distintas concepciones educacionales, ya que pondría en peligro la necesaria coherencia de la acción educativa.

Sin embargo, sí sería posible y quizá necesario para el avance científico probar distintas alternativas de elementos específicos en el contexto de un modelo de evaluación básicamente común. Ciertamente que la selección de estos elementos que, sin amenazar la unidad y coherencia del Centro, pudieran arrojar alguna luz sobre el efecto de la evaluación en el rendimiento escolar, exige hipótesis apoyadas en la experiencia o en teorías acerca de los modos y vías de influencia de la evaluación en los procesos y productos educacionales. En nuestra opinión, lo que ya sabemos acerca de las relaciones entre los distintos elementos de un modelo evaluativo y el comportamiento discente de los alumnos permitiría avanzar significativamente en esta línea. Creemos que pudieran ser objeto de contraste, hipótesis sobre la influencia en el rendimiento de elementos evaluativos como los siguientes:

- Énfasis en extensión de conocimientos versus énfasis en nivel de dificultad de las tareas.
- Énfasis en diversos tipos de pruebas en la evaluación.
- Énfasis en distintos calendarios y periodicidad de aplicación de las pruebas.
- Énfasis en el nivel y modo de información previa acerca de la evaluación.
- Énfasis entre distintos modos de informar a los alumnos de los resultados de la evaluación.

Cada prueba en un centro podría considerarse como una replicación del mismo experimento, lo que permitiría la generalización de resultados a contextos institucionales diferenciados.

REFERENCIAS

- CROOKS, T. J. (1988). The Impact of Classroom Evaluation Practices on Students. *Review of Educational Research* 58, 4, 438-481.
- DE LA ORDEN, A. (1969). Evaluación del rendimiento educativo y la calidad de la enseñanza. *Revista de Educación*, XVIII, nº 206.
- DE LA ORDEN, A. (1981). *Evaluación del aprendizaje y calidad de la educación* En Varios, *La calidad de la educación* Madrid: CSIC.
- DE LA ORDEN, A. (1982). *La evaluación educativa. Concepto, características, funciones*. Buenos Aires: Docencia-Proyecto CINAIE.
- DE LA ORDEN, A. (1983). La investigación sobre la evaluación educativa. *Revista de Investigación Educativa*, 2, 240-258.
- DE LA ORDEN, A. (1988). La calidad de la educación. *Bordón*, 42,2,149-161.
- FELDHUSEN et al. (1977). *Is a lack of instructional validity contributing to the decline of achievement test scores?* En L. Lipsitz (Ed.) *The test scores decline: meaning and uses*. Englewood Cliffs. Educational technology Publications.
- GARCÍA YAGÜE, J. y PALOMINO LÓPEZ, A. (1976). *Batería Factorial de la Inteligencia AMD*. Valladolid: Miñón.
- LÁZARO MARTÍNEZ, A. (1989). *Pruebas de comprensión lectora*. Madrid: TEA.
- OLTMAN, RASKIN y WITKIN (1981). *Test de figuras enmascaradas (GEFT)*. Madrid: TEA.
- SHOEMAKER, D. M. (1975). Toward a framework of achievement testing. *Review of Educational Research*, 45, 127-147.