

ANÁLISIS DE LAS PRUEBAS FINALES DE E.G.B.*

por ARTURO DE LA ORDEN y ANTONIO GONZÁLEZ SOLER
Universidad Complutense de Madrid

Consideraciones generales y objetivos de la investigación

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, 1981) formulaba esta tesis 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 misma 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

* Este artículo se apoya fundamentalmente en los resultados de un trabajo de investigación que, bajo el epígrafe "Estudio comparativo de las pruebas finales de E.G.B.", se llevó a cabo durante los años 1978-1979, en el Instituto de Ciencias de la Educación de la Universidad Complutense de Madrid, por un equipo formado por Antonio González Soler, Jaime Acebrón Orozco, Eduardo Soler Fierrez, Manuel Vázquez Fernández, Bonifacio Llorente Hernández e Ignacio González Carrasco, y dirigido por Arturo de la Orden.

El proyecto, incluido en el VIII Plan Nacional de Investigaciones Educativas, fue financiado por el INCIE.

enseñanza en función de los criterios de tales pruebas formuladas o no expresamente. Lo que 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 en 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. Analizando este tipo de exámenes y sus posibilidades de reforma, Eggleston y Kerr (1969) aluden a los inconvenientes de "introducir otro examen que *determinaría en gran medida los objetivos, el contenido y los métodos de enseñanza*". En nuestro contexto son conocidos los efectos sobre los objetivos de la enseñanza media de los antiguos exámenes de estado y las reválidas, 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 constatar que como las actuales pruebas de acceso a la Universidad han dado un giro radical a los objetivos del curso de orientación, hasta el punto de que la orientación sólo aparece en el nombre, lo cual no dejar 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 como lo enseñan, los contenidos y los métodos, en otras palabras, el producto y el proceso de la educación. Este es, a nuestro 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, como he afirmado en otro lugar (De la Orden, 1981), a un sistema de coherencias en el proceso total de la educación, la virtualidad de la evaluación como estímulo o freno de la calidad educativa, dependerá enteramente de las características del esquema evaluador. Si los criterios y modos de evaluación, que determinan los objetivos reales y en última instancia el producto de la enseñanza y el aprendizaje son coherentes con los objetivos formalmente establecidos —o sobre los que existe un consenso implícito— 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, los criterios y modos de evaluación movilizan los procesos educativos hacia objetivos y resultados no coherentes con las metas formalmente establecidas o implícitamente aceptadas, ni en consecuencia, con los fines generales de la educación y el sistema axiológico en que se sustentan, la evaluación se constituye en el mayor obstáculo a una educación de calidad. Este sería el caso, por ejemplo, de un sistema de exámenes cuya característica esencial fuera exigir la reproducción de información fáctica, contribuyendo así a fijar como objetivo real de la enseñanza y el aprendizaje la asimilación, reconocimiento y evocación de datos más o menos relevantes. Estos objetivos reales canalizarán casi necesariamente las energías y actividad de alumnos y profesores hacia un aprendizaje memorístico y fomentarán los métodos didácticos verbalistas y repetitivos y el abuso del texto entendido como catecismo. Si las metas y objetivos previstos para la enseñanza no coinciden con la memorización de datos e información verbal —supongamos que fueran el desarrollo de la capacidad crítica frente a los hechos y de la capacidad para usar la información aplicándola a la resolución de problemas— es evidente que tal sistema de exámenes constituye un formidable obstáculo a la calidad de la educación. Si, por el contrario, el sistema de exámenes se caracterizara por exigir de los alumnos análisis de situaciones supuestas y reales, valoración de información y aplicación de sus conocimientos, determinaría unos objetivos reales coincidentes con los previstos y movilizaría la actividad docente y discente hacia una didáctica que obligaría a los alumnos a pensar por sí mismos, a emitir y justificar juicios de valor, a resolver problemas con sentido para ellos, etc. En otras palabras, un tal sistema de exámenes sería una poderosa palanca para impulsar la calidad de la educación.

Partiendo, pues, del supuesto de que las características de la evaluación educativa condicionan en gran medida lo que el profesor enseña y lo que el alumno aprende, así como la forma en que ambos procesos se desarrollan, a través de esta investigación se pretende analizar las pruebas de 8.º de E.G.B. (Lengua y Matemáticas) que confeccionan y aplican los Colegios de Madrid y su provincia, bajo la supervisión de la Inspección Técnica de Educación.

Con este análisis perseguimos 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 Básica a nuestros alumnos.

A efectos de análisis se distinguen dos grandes apartados:

a) Contenido de las pruebas.

— Adecuación de los contenidos

— Nivel de los contenidos

— Capacidades intelectuales que miden y el 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

En último término, lo que se pretende averiguar con esta investigación se puede expresar en estas preguntas fundamentales:

— ¿Qué se exige realmente a los alumnos en Lengua y Matemáticas al finalizar el período de Educación General Básica y obtener el título de Graduado Escolar?

— ¿Existe unanimidad entre los Colegios sobre lo que significa en términos de conocimientos, capacidades, actitudes, el título de Graduado Escolar, como sanción social de la Educación General Básica? En caso negativo, ¿cuál es el grado de concordancia a este respecto?

— ¿Se usan instrumentos adecuados para determinar qué alumnos responden convenientemente a ese concepto (o conceptos) de Educación General Básica y Graduado Escolar?

La opinión generalizada en una parte de nuestra sociedad de que la implantación de la Educación General Básica ha supuesto un descenso del pretendido nivel intelectual del antiguo Bachillerato Elemental, constituye una afirmación ambigua y no contrastada y que, sin embargo, puede afectar negativamente al desarrollo educativo normal de los alumnos si, entre padres y profesores, alcanzara altos niveles de difusión. El análisis de lo que se exige a los alumnos al finalizar el período de Educación General Básica puede ayudar a esclarecer el carácter, sentido y nivel de esta etapa fundamental de la educación de los españoles.

Al mismo tiempo, la determinación objetiva de las características de las pruebas más importantes a que son sometidos los alumnos durante este período, surge de la necesidad de clarificar hasta donde ello sea posible, la controversia sobre la evaluación que siguió a la modificación de esta faceta educativa como consecuencia de la puesta en marcha de la Ley de Educación.

En el momento de transición político-social a que se halla sometida la sociedad española, parece lógico suponer que uno de los sectores prioritarios en la reforma general será la educación. Por tanto, los resultados del presente estudio pueden constituir una base objetiva para abordar con mayor eficacia las, al parecer, inevitables modificaciones en la estructura y contenido de los programas de E.G.B. y en los modos de evaluación prevalentes.

Diseño de la investigación

1. 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 E.G.B. 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. La gama de variables dependientes es la siguiente:
 - 1) Adecuación de los contenidos

- 2) Nivel de dificultad de la prueba
- 3) Capacidades que pretenden medir
- 4) Peso relativo de cada capacidad dentro de la prueba
- 5) Adecuación del tipo de ítem a la capacidad que pretende medir
- 6) Adecuación de construcción de la prueba
- 7) Extensión de la prueba

El número de posibles hipótesis de influencia de cada variable independiente sobre las variables dependientes es muy alto y su comprobación rebasaría las pretensiones y posibilidades y recursos de la presente investigación. Por ello, hemos decidido reducir el número de variables dependientes, siguiendo un criterio de mayor relevancia y posibilidad de cuantificación y análisis estadístico de los datos. No obstante, tendremos en cuenta en nuestro estudio el resto de las variables.

Las variables dependientes seleccionadas son las siguientes:

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

Como consecuencia de esta reducción se han generado las siguientes hipótesis de nulidad:

- 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 independientes ni el control de aquellas otras variables intervinientes ajenas al interés de los investigadores.

2. 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 su 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.

3. Definición de los estratos

Urbano. Dentro de esta categoría están comprendidos los centros enclavados en el casco de Madrid capital. Queda excluido el llamado "cinturón industrial" y "ciudades dormitorio".

Rural. El establecimiento de los límites entre las categorías rural, semiurbano y urbano ofrece no pocas dificultades a los investigadores. En nuestro caso, el problema presenta una característica especial: la de que los centros enclavados en dichas localidades deben tener como mínimo ocho unidades escolares, es decir, ser Centros completos de E.G.B. Este factor reduce considerablemente el número de localidades rurales objeto del estudio. Por ello, nos hemos visto obligados a incluir en este estrato localidades que podrían en puridad categorizarse como semirurales. El estrato ha quedado definido así: Colegios de más de ocho unidades enclavados en localidades no industrializadas, ni residenciales de menos de 5.000 habitantes.

Estatal. Esta categoría comprende los Centros creados, mantenidos y regidos por el Estado. No se incluyen los denominados Centros de Patronato.

Privados. Se incluyen en esta categoría los Centros creados, mantenidos y regidos por particulares, aunque estén subvencionados por el Estado.

Nivel socio-económico. La conceptualización de los tres niveles identificados (alto, medio, bajo), se ofrece también dificultosa por la variedad de indicadores que se pueden conjugar: profesión, nivel de educación, ingresos anuales, etc. Las obvias dificultades de identificar el nivel socioeconómico de cada centro mediante el análisis pormenorizado de dichos indicadores con respecto a cada familia, nos ha obligado a recurrir a un procedimiento más asequible a las posibilidades económicas, de personal y de tiempo, de este proyecto. Hemos clasificado los distritos de Madrid en las tres categorías anteriormente expresadas usando a tal fin los datos de la Encuesta de Origen y Destino, realizada por COPLACO en 1973, que utiliza el nivel de renta y la profesión como principales indicadores (citada por Olivera Poll, 1976). Además, hemos usado un nuevo indicador: la condición de *Subvencionado o no*, aplicable a los Centros Privados, dado que en la concesión de subvención a un Centro, se utiliza como criterio fundamental el nivel socio-económico de los alumnos que acoge. Este indicador se ha mostrado especialmente relevante en las zonas o distritos de nivel alto, causando la eliminación de los centros

subvencionados de las mismas de la categoría o nivel privado-alto, por la circunstancia expresada en el párrafo anterior.

Nivel socio-económico alto (Sólo se aplica a Centros Privados, y² que como hicimos constar anteriormente, nos encontramos centros estatales de nivel económico comparable a los privados en esta categoría). Centros privados no Subvencionados enclavados en los distritos Moncloa, Chamberí, Chamartín, Salamanca y Retiro. Centros privados No subvencionados enclavados en zonas residenciales en la periferia y que se nutren de alumnos mediante transporte, la Moraleja, la Florida, Pozuelo, etc.

Nivel socio-económico medio. Esta categoría incluye Centros estatales y privados subvencionados al 100 por 100 enclavados en los distritos mencionados en el apartado anterior, y también Centros nacionales y privados (subvencionados o no), enclavados en los distritos Centro, Arganzuela, Tetuán, Fuencarral, Latina, Moratalaz, Ciudad Lineal y Hortaleza.

Nivel socio-económico bajo. Comprende Centros estatales y privados, Subvencionados o No, enclavados en los distritos siguientes: Carabanchel, Villaverde, Mediodía, Vallecas y San Blas.

Una vez definidos los estratos e identificados los Centros de más de ocho unidades pertenecientes a cada uno de ellos, se extrajo por procedimientos aleatorios (Tabla de números aleatorios), una muestra que consta de:

- 30 Colegios rurales estatales
- 15 Colegios urbanos estatales, clase media
- 15 Colegios urbanos estatales, clase baja
- 15 Colegios urbanos privados, clase alta
- 15 Colegios urbanos privados, clase media
- 15 Colegios urbanos privados, clase baja

4. *Carácter de los datos y técnicas de medida*

4.1. *Datos relativos a la primera variable dependiente: Adecuación y relevancia de los contenidos en las pruebas de Lengua y Matemáticas:*

Los datos relativos a esta variable se han concretado en las puntuaciones asignadas a las pruebas de los distintos centros seleccionados, como resultado de la aplicación a las mismas de una escala construida al efecto.

4.1.1. *Escala aplicada a las pruebas de Lengua Española*

Punto de partida:

- Bases de programación por áreas educativas del MEC.
- Cómputo de líneas efectuado sobre los cinco libros de texto con más implantación en el mercado.

Problemas planteados

- Las Orientaciones del MEC son tan sólo un marco de sugerencias que posibilitan múltiples interpretaciones.

— Existencia de una gran variedad de modelos de descripción gramatical que dificultan la posibilidad de puntos de contacto uniformes entre los distintos libros de texto.

— Los libros de texto reflejan el panorama un tanto caótico de la ciencia del lenguaje en la actualidad: el libro de texto está concebido en términos de enseñanza, no de aprendizaje.

Soluciones adoptadas:

— Partir de conceptos muy generales de la ciencia del lenguaje que permitan una relativa homogeneización de conceptos.

— Establecer las líneas maestras de lo que debe ser la enseñanza de la lengua en EGB en cuanto a programación se refiere. Se establecen tres aspectos vertebradores de la enseñanza del lenguaje:

- Aspecto de desarrollo de la competencia lingüística que los alumnos poseen como hablantes nativos.

- Aspecto de estudio reflexivo de la propia competencia lingüística.

- Aspecto de aproximación a la lengua literaria y a sus manifestaciones históricas.

— La puntuación de la prueba se entenderá referida al peso específico de cada uno de los aspectos en los libros de texto: equilibrio de la prueba.

Consideraciones tenidas en cuenta en la elaboración y aplicación de la escala

— El número de líneas de cada elemento del programa sólo puede relacionarse con los de su mismo aspecto y jerarquía.

— Siguiendo una necesidad de homogeneización no se detalla mucho el número de subaspectos.

— En la puntuación de la prueba se tiene en cuenta la suma de grados de representatividad parcial.

— La puntuación total de la prueba se obtiene de la suma de la de cada uno de los aspectos parciales.

La calificación de cada aspecto parcial se realiza de acuerdo con la escala siguiente:

1.^º de 0 a 16

2.^º de 0 a 10

3.^º de 0 a 11

— Cada subaspecto tiene un peso relativo proporcional a su relevancia.

— Existen subaspectos no considerados en los libros de texto.

Escala

A (16 puntos). *Desarrollo de la competencia lingüística que los alumnos poseen como hablantes nativos*

— Comprensión y expresión escrita

- Análisis y síntesis de textos escritos 6

- Técnicas y ejercicios de expresión escrita 4

- Desarrollo del vocabulario 1

- Ortografía 2

ANALISIS DE LAS PRUEBAS FINALES DE E.G.B. 15

- Comprensión y expresión oral
- Pronunciación correcta 1
- Técnicas y ejercicios de expresión oral 2

B (11 puntos). *Estudio reflexivo de la propia competencia lingüística*

- Principios generales de lingüística 0,5
- Fonología 1
- Morfosintaxis
- Análisis morfológico 2
- Estudio de la oración simple 1
- La oración compuesta 5
- Lexicología y semántica 1,5

C (10 puntos). *Aproximación a la lengua literaria y a sus manifestaciones históricas*

- La creación literaria. Recursos expresivos. La métrica 1
- Los géneros literarios 1
- La obra literaria en su contenido histórico-social 8

37

Puntuación máxima de la prueba = A + B + C = 37 puntos

4.1.2. *Escala aplicada a las pruebas de Matemáticas*

Partiendo de los mismos supuestos y aplicando los mismos principios, se elaboró la siguiente escala para determinar y cuantificar la adecuación y relevancia de los contenidos de las pruebas de Matemáticas:

1 (1,50 puntos). *Conjuntos*

- 1.1. Operaciones en P (A) 0,50
- 1.2. Relaciones 0,50
- 1.3. Estructuras 0,50

2 (0,50 puntos). *Múltiplos y Divisores*

- 2.1. Criterios de divisibilidad 0,25
- 2.2. M.C.M. y M.C.D. 0,25

3 (7 puntos). *El N° Q (operaciones)* 7

4 (1 punto). *Potenciación (operaciones)* 1

5 (3,50 puntos). <i>Polinomios</i>	0,50
5.1. Monomios (operaciones)	2,50
5.2. Polinomios (operaciones)	0,50
5.3. Expresiones notables	
6 (1 punto). <i>Radicación (operaciones)</i>	1
7 (2,50 puntos). <i>Funciones</i>	
7.1. De variable Q.	1,25
7.2. Polinómica	1,25
8 (7,50 puntos). <i>Ecuaciones</i>	
8.1. De primer grado	2
8.2. Sistemas	1
8.3. De segundo grado	2,25
8.4. Sistemas	2
8.5. Parábola e hipérbola	0,25
9 (0,25 puntos). <i>Proporcionalidad de segmentos</i>	0,25
10 (0,50 puntos). <i>Semejanzas</i>	
10.1. Concepto	0,20
10.2. De triángulos	0,20
10.3. Razón áreas figuras	0,10
11 (0,25 puntos). <i>Relaciones métricas en el triángulo rectángulo</i>	0,25
12 (0,25 puntos). <i>Ángulos en la circunferencia</i>	0,25
13 (0,25 puntos). <i>Nociones de Trigonometría</i>	0,25
	26,00

Puntuación máxima de la prueba = 26 puntos

- Cuando en un ítem se incluyan varios aspectos o subaspectos diferentes, se puntuará con la media de las puntuaciones correspondientes a cada uno de tales aspectos o subaspectos.
- Todos los ítems que traten del mismo aspecto o subaspecto se puntuarán en conjunto una sola vez.
- Si en un ítem entran dos o más aspectos bien diferenciados se puntuarán por separado.
- Si un ítem se encuentra en la prueba sin puntuar, es porque se le ha incluido anteriormente en algún aspecto o subaspecto idéntico.

4.2. *Datos relativos a la segunda variable dependiente: Peso relativo de las capacidades medidas en las pruebas de Lengua y Matemáticas*

Los datos relativos a esta variable se han concretado en porcentaje o proporción de ítems destinados a medir cada capacidad identificada, respecto al total de ítems incluidos en la prueba.

Para la determinación de las distintas capacidades intelectuales a considerar se recurrió a las seis categorías del dominio cognoscitivo señaladas por B. S. Bloom y otros (1972).

Por tanto, cada prueba queda calificada en esta variable por seis porcentajes, correspondientes a la proporción de sus ítems dedicados respectivamente a medir las siguientes capacidades, según son definidas por Bloom y sus colaboradores:

- Conocimiento
- Comprensión
- Aplicación
- Análisis
- Síntesis
- Evaluación

En ningún caso se analizaron las subcategorías generales.

4.3. *Datos relativos a la tercera variable dependiente: Características formales de las pruebas*

Los datos relativos a esta variable se han concretado en las puntuaciones asignadas a las pruebas de los distintos Centros seleccionados, como resultado de la aplicación a los mismos de dos escalas (una para análisis de los ítems y otra que considera globalmente la prueba) construidas al efecto. Los criterios seguidos en la construcción de estas escalas se apoyan fundamentalmente en las normas de construcción de instrumentos de evaluación normalmente incluidas en los manuales de psicometría más utilizados y que se citan en la bibliografía.

4.3.1. *Escala aplicada a las pruebas de Lengua Española*

Las pruebas de lenguaje están construidas fundamentalmente por dos clases de cuestiones: ítems de respuesta objetiva y pruebas de ensayo.

Los ítems objetivos sobre los que están construidas son de los siguientes tipos:

- Verdadero-falso
- Elección múltiple (de 3 ó 4 acepciones)
- Correspondencia
- Completar frases

Además, hay ítems de respuesta abierta, cuya construcción está apoyada generalmente en un texto literario, y ejercicios de análisis sintáctico y morfológico.

Su estructura se ha calificado desde dos puntos de vista:

a) En primer lugar, se ha tenido en cuenta la redacción de los ítems o ejercicios, analizando cada uno de ellos para detectar los errores que pudieran invalidarlos como instrumentos de medición.

A estos efectos se elaboró la siguiente escala de calificación para cada una de las cuestiones consideradas:

 ITEMS DE VERDADERO-FALSO

Errores más corrientes	Puntos a deducir
— Enunciados verdaderos y falsos en parte	6
— Uso de dobles negaciones	4
— Respuestas correctas más largas o mejor redactadas que las falsas	3
— Uso de determinadores específicos	3
— Enunciados demasiado largos (más de 25 palabras)	2
— Elemento principal del enunciado situado en una oración subordinada	1

 ITEMS DE ELECCION MULTIPLE

Errores más corrientes	Puntos a deducir
— Más de una respuesta correcta en items de una sola respuesta	6
— Respuestas gramaticalmente incompatibles con el interrogante	4
— Elementos de distracción no plausibles	3
— Respuestas correctas más largas o mejor redactadas que las falsas	3
— Uso de determinadores específicos	3
— Respuestas demasiado largas, con repeticiones innecesarias	2

 ITEMS DE CORRESPONDENCIA

Errores más corrientes	Puntos a deducir
— Material inconsistente o heterogéneo en una misma columna	6
— Pocos elementos por ítem (menos de 5)	5
— Muchos elementos por ítem (más de 10)	3
— Usos de determinadores específicos (elementos idénticos, enunciados ilógicos, pistas gramaticales)	3
— Columna de opciones no ordenada según algún criterio sistemático	2
— Mismo número de preguntas en ambas columnas	1

 ITEMS DE COMPLETAR FRASES

Errores más corrientes	Puntos a deducir
— Falta de objetividad (más de una respuesta posible)	6
— Omisión de palabras triviales	6
— Demasiados espacios en blanco en un ítem	5
— Frases ambiguas	4
— Espacios en blanco al principio de la frase o irregularmente distribuidos	3
— Espacios en blanco de longitud variable	1

 ITEMS DE ENSAYO

Errores más corrientes	Puntos a deducir
— Enunciado mal redactado, de tal forma que la tarea a realizar por el alumno no queda definida claramente (preguntas ambiguas o imprecisas)	6
— Falta de instrucciones generales para contestar el ítem	3
— Enunciado cuya respuesta depende solamente de la capacidad para reproducir información (suelen ser así todos los ítems que empiecen por "qué", "quién", "cuándo" y "enumere")	1

Teniendo en cuenta estas escalas, cada ítem se ha calificado de la siguiente manera:

1. A cada ítem se le han asignado de entrada 6 puntos.
2. Teniendo en cuenta el tipo de ítem, se le han deducido los puntos correspondientes al error descubierto en redacción, según la clave considerada.
3. Por último, se le ha asignado al ítem la calificación correspondiente según los apartados 1 y 2.

La calificación del ítem está, pues, entre 0 y 6, siendo tanto mejor cuanto mayor sea la puntuación obtenida.

Para la calificación de cada prueba se ha obtenido la media de los puntos asignados a los ítems de la misma. La puntuación de la prueba oscila entre 0 y 6.

- b) Por otra parte, se ha examinado cada una de las pruebas consideradas en su

conjunto, tratando de determinar si su estructura era la más correcta respecto a extensión, variedad y distribución de los ítems.

4.3.2. Escalas aplicadas a las pruebas de Matemáticas

Dadas sus características, a las pruebas de Matemáticas se aplicó la siguiente escala:

~ ITEMS DE MATEMATICAS ~	
Errores más corrientes	Puntos a deducir
I. Enunciado o planteamiento con errores	6
II. Carece del suficiente número de datos para resolver el ejercicio	6
III. El enunciado está mal redactado o utiliza un lenguaje (verbal, simbólico o gráfico) poco preciso y exacto, de tal forma que la tarea a realizar por el alumno no queda definida claramente	3
IV. El ítem tiene más de una pregunta	2
V. Falta de instrucciones, de tal modo que el alumno debe deducir lo que tiene que hacer por el planteamiento del ejercicio	1

La forma de calificar cada uno de los ítems o ejercicios y el baremo al que se someterán las pruebas en su conjunto son los mismos que se han descrito en el apartado correspondiente a las pruebas de lenguaje.

RESULTADOS DEL ESTUDIO Y ANALISIS DE LOS DATOS

1. Resultados

A continuación se presentan los datos (medias, varianzas, proporciones, porcentajes) que definen características correspondientes a los diversos estratos de la muestra y a la muestra en su conjunto.

1.1. Datos obtenidos en la primera variable: Adecuación y relevancia de los contenidos de las pruebas

1.1.1. Pruebas de Lengua - 8.º de EGB

	CENTROS PRIVADOS				Extracción Social												
					Alta				Media				Baja				
	Total	A	B	C	Total	A	B	C	Total	A	B	C	Total	A	B	C	
N	45	45	45	45	15	15	15	15	15	15	15	15	15	15	15	15	15
\bar{X}	21,71	6,91	6,67	8,31	22,80	7,27	7,67	9,07	21,16	7,48	5,63	8,13	21,16	6,06	6,70	8,40	
S ²	36,79	14,99	9,08	5,48	20,79	15,20	2,85	6,98	44,70	12,10	12,90	5,41	46,80	18,60	10,60	4,69	
	CENTROS PUBLICOS URBANOS				Extracción Social												
					Media				Baja								
	Total	A	B	C	Total	A	B	C	Total	A	B	C					
N	30	30	30	30	15	15	15	15	15	15	15	15					
\bar{X}	25,66	8,83	7,70	9,13	27,13	10,27	7,47	9,47	24,13	7,4	7,93	8,80					
S ²	17,28	3,6	1,61	1,68	17,12	10,10	5,77	0,55	15,05	14,26	3,85	5,31					
	CENTROS PUBLICOS				CENTROS PUBLICOS RURALES				TODOS LOS CENTROS								
	Total	A	B	C	Total	A	B	C	Total	A	B	C					
	N	60	60	60	60	30	30	30	30	105	105	105	105				
\bar{X}	24,08	7,37	8,06	8,68	22,53	5,93	8,27	8,23	22,90	7,38	7,14	8,38					
S ²	33,38	18,75	3,95	5,40	45,59	19,93	3,04	7,67	35,36	15,89	7,90	6,50					

1.1.2. Pruebas de Matemáticas - 8.º de EGB

	CENTROS PRIVADOS	Extracción Social		
		Alta	Media	Baja
N	45	15	15	15
\bar{X}	14,21	14,83	13,90	13,76
S ²	16,27	16,69	22,84	12,74
	CENTROS PUBLICOS URBANOS	Extracción Social		
		Media	Baja	
N	30	15	15	
\bar{X}	13,73	12,80	14,66	
S ²	20,98	22,06	18,17	
CENTROS PUBLICOS				
N	60			
\bar{X}	13,52			
S ²	18,63			
CENTROS PUBLICOS RURALES				
N	30			
\bar{X}	13,31			
S ²	16,22			
TODOS LOS CENTROS				
N	105			
\bar{X}	13,87			
S ²	18,02			

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

1.2.1. Pruebas de Lengua - 8.º de EGB

CENTROS PRIVADOS	Capacidades medidas (% de items dedicadas a cada una)					
	Conocimiento	Comprensión	Aplicación	Análisis	Síntesis	Evaluación
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	57	15	9	18		1
CENTROS PUBLICOS						
— 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

1.2.2. Pruebas de Matemáticas - 8.º de EGB

CENTROS PRIVADOS	Capacidades medidas (% de items dedicados a cada una)					
	Conoci- miento	Com- prensión	Aplica- ción	Análisis	Síntesis	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 PUBLICOS						
— Urbanos:						
Extracción Social Media	76	11	13			
Extracción Social Baja	66	13	21			
— Rurales:	62	21	17			
Total Centros Públicos Urbanos	71	12	17			
Total Centros Públicos	67	16	17			
Total Centros Públicos y Privados	65	19	16		0	

1.3. Datos obtenidos en la tercera variable: Características formales de las pruebas

— Porcentaje de los distintos tipos de ítems utilizados en las pruebas elaboradas por los diferentes grupos de Centros —Pruebas de Lengua— 8.º de E.G.B.

Centros	Tipos de Ítems					Total Ítems
	V—F	E.M	C	C.F	E	
1. Centros Privados	0,66	1,49	2,32	0,33	95,36	603
—Extracción Social Alta	1,76	3,08	3,96	—	91,18	227
—Extracción Social Media	—	—	0,55	—	99,45	180
—Extracción Social Baja	—	1,02	2,04	1,02	96,42	196
2. Centros Públicos	—	5,53	2,10	1,44	90,92	904
—Centros Públicos Urbanos	—	10,90	2,32	1,85	84,91	431
—Extracción Social Media	—	12,62	0,97	3,40	83,01	206
—Extracción Social Baja	—	9,33	3,56	0,44	86,67	225
—Centros Públicos Rurales	—	0,63	1,90	1,06	96,40	475
3. Centros Públicos y Privados	0,26	3,91	2,19	0,99	92,70	1.507

— Porcentaje de los distintos tipos de ítems utilizados en las pruebas elaboradas por los diferentes grupos de Centros —Pruebas de Matemáticas— 8.º de E.G.B.

Centros	Tipos de Ítems				Total Ítems
	Teoría y definiciones	Ejercicios de desarrollo	Demostraciones	Problemas	
1. Centros Privados	12,63	61,32	4,16	21,88	649
—Extracción Social Alta	24,36	46,19	6,09	23,35	197
—Extracción Social Media	9,27	72,59	2,32	15,83	259
—Extracción Social Baja	5,18	61,66	4,66	28,50	193
2. Centros Públicos	11,90	61,91	2,25	23,93	932
—Centros Públicos Urbanos	11,49	57,88	1,80	28,83	444
●Extracción Social Media	13,82	60,83	1,84	23,50	217
●Extracción Social Baja	9,25	55,07	1,76	33,92	227
—Centros Públicos Rurales	12,29	65,57	2,66	19,47	488
3. Centros Públicos y Privados	12,20	61,67	3,04	23,09	1.581

— *Adecuación de la construcción de la prueba*

— *Puntuaciones medias de las pruebas por Centros y grupos de Centros — Lengua—
8.º de E.G.B.*

CENTROS PRIVADOS	N.º total de ítems	\bar{X} total
Extracción Social Alta	227	5,45
Extracción Social Media	180	4,98
Extracción Social Baja	196	4,97
Total Centros Privados	603	5,16
CENTROS PUBLICOS		
— Urbanos:		
Extracción Social Media	206	5,08
Extracción Social Baja	225	5,37
— Rurales:	473	5,14
Total Centros Públicos Urbanos	431	5,23
Total Centros Públicos	904	5,18
Total Centros Públicos y Privados	1.507	5,17

*Puntuaciones medias de las pruebas por Centros y grupos de Centros
Matemáticas - 8.º de EGB*

CENTROS PRIVADOS	N.º total de ítems	\bar{X} total
Extracción Social Alta	197	5,50
Extracción Social Media	259	5,30
Extracción Social Baja	193	5,50
Total Centros Privados	649	5,40
CENTROS PUBLICOS		
— Urbanos:		
Extracción Social Media	217	5,30
Extracción Social Baja	227	5,40
— Rurales:	488	5,50
Total Centros Públicos Urbanos	444	5,30
Total Centros Públicos	932	5,50
Total Centros Públicos y Privados	1.581	5,40

2. Análisis de los datos

2.1. Análisis realizados en la primera variable: Adecuación y relevancia de los contenidos de las pruebas

— Análisis de la varianza de las pruebas de Lengua de 8.^º para los Centros Públicos y Privados que atienden alumnos de clase social media y baja. (Puntuación total de la prueba).

Fuentes de variación	Suma de cuadrados	g.l.	Varianza estimada	F
C Entre Centros	299,27	1	299,27	12,40●●
E Entre extracción S.	33,75	1	33,75	1,39
Interacción C × E	33,74	1	33,74	1,39
Dentro de grupos	1.350,64	56	24,12	
Total	1.717,40	59		

●● Significativa al 0,01

— Análisis de la varianza de las pruebas de Lengua de 8.^º para los Centros Públicos y Privados que atienden alumnos de clase social media y baja. (Puntuaciones en el apartado A).

Fuentes de variación	Suma de cuadrados	g.l.	Varianza estimada	F
C Entre Centros	66,15	1	66,15	4,81*
E Entre extracción S.	66,15	1	66,15	4,81*
Interacción C × E	8,82		8,82	< 0
Dentro de grupos	771,06		13,76	
Total	912,18	59		

• Significativa al 0,05

— *Análisis de la varianza de las pruebas de Lengua de 8.º para los Centros Públicos y Privados que atienden alumnos de clase social media y baja. (Puntuación en el apartado B).*

Fuentes de variación	Suma de cuadrados	g.l.	Varianza estimada	F
C Entre Centros	35,26	1	35,26	4,26●
E Entre extracción S.	8,82	1	8,82	1,07
Interacción C × E	1,35	1	1,35	< 0
Dentro de grupos	463,30	56	8,27	
Total	508,73	59		

● Significativa al 0,05

— *Análisis de la varianza de las pruebas de Lengua de 8.º para los Centros Públicos y Privados que atienden alumnos de clase social media y baja. (Puntuación en el apartado C).*

Fuentes de variación	Suma de cuadrados	g.l.	Varianza estimada	F
C Entre Centros	11,26	1	11,26	2,82
E Entre extracción S.	0,60	1	0,60	< 0
Interacción C × E	3,27	1	3,27	< 0
Dentro de grupos	223,47	56	3,99	
Total	238,60	59		

— *Análisis de la varianza de las pruebas de Matemáticas de 8.º para los Centros Públicos y Privados que atienden alumnos de clase social media y baja.*

Fuentes de variación	Suma de cuadrados	g.l.	Varianza estimada	F
C Entre Centros	0,81	1	0,81	< 0
E Entre extracción S.	21,6	1	21,60	< 0
Interacción C × E	6,70	1	6,70	< 0
Dentro de grupos	1.408,10	56	25,14	
Total	1.437,21	59		

— Prueba de t para el contraste de la diferencia entre las medias de las pruebas de Lengua de 8.º para los Centros Públicos Urbanos y Rurales. (Puntuación total de la prueba).

Centros	N	\bar{X}	S ²
Urbanos	30	25,66	17,18
Rurales	30	22,53	43,73

$$t = 2,18^*$$

$$g.l. = 58$$

* Significativa al 0,05

— Prueba de t para el contraste de la diferencia entre las medias de las pruebas de Lengua de 8.º para los Centros Públicos Urbanos y Rurales. (Puntuación en el apartado A).

Centros	N	\bar{X}	S ²
Urbanos	30	8,83	13,46
Rurales	30	5,93	19,30

$$t = 2,78^{**}$$

$$g.l. = 58$$

** Significativa al 0,01

— Prueba de t para el contraste de la diferencia entre las medias de las pruebas de Lengua de 8.º para los Centros Públicos Urbanos y Rurales. (Puntuación en el apartado B).

Centros	N	\bar{X}	S ²
Urbanos	30	7,70	4,54
Rurales	30	8,43	3,00

$$t = 1,46$$

$$g.l. = 58$$

— Prueba de t para el contraste de la diferencia entre las medias de las pruebas de Lengua de 8.º para los Centros Públicos Urbanos y Rurales. (Puntuación en el apartado C).

Centros	N	\bar{X}	S ²
Urbanos	30	9,13	2,91
Rurales	30	8,23	1,40

$$t = 2,36^*$$

$$g.l. = 58$$

* Significativa al 0,05

— Prueba de *t* para el contraste de la diferencia entre las medias de las pruebas de Matemáticas de 8.º de los Centros Públicos Urbanos y Rurales.

Centros	N	\bar{X}	S ²
Urbanos	30	13,73	20,99
Rurales	30	13,31	16,22

$$t = < 0$$

— Análisis de la varianza de las pruebas de Lengua de 8.º para los Centros Privados de clase alta, media y baja. (Puntuación total de la prueba).

Fuentes de variación	Suma de cuadrados	g.l.	Varianza estimada	F
Entre grupos	26,68	2	13,34	<0
En grupos	1.155,07	42	27,50	
Total	1.181,75	44		

— Análisis de la varianza de las pruebas de Lengua de 8.º para los Centros Privados de clase alta, media y baja. (Puntuación en el apartado A).

Fuentes de variación	Suma de cuadrados	g.l.	Varianza estimada	F
Entre grupos	16,17	2	8,08	< 0
En grupos	522,47	42	12,44	
Total	538,64	44		

— Análisis de la varianza de las pruebas de Lengua de 8.º para los Centros Privados de clase alta, media y baja. (Puntuación en el apartado B).

Fuentes de variación	Suma de cuadrados	g.l.	Varianza estimada	F
Entre grupos	31,03	2	15,52	1,76
En grupos	368,47	42	8,77	
Total	399,50	44		

— *Análisis de la varianza de las pruebas de Lengua de 8.º para los Centros Privados de clase alta, media y baja. (Puntuación en el apartado C).*

Fuentes de variación	Suma de cuadrados	g.l.	Varianza estimada	F
Entre grupos	2,13	2	1,07	<0
En grupos	239,07	42	5,69	
Total	241,20	44		

— *Análisis de la varianza de las pruebas de Matemáticas de 8.º para los Centros Privados que atienden alumnos de clase alta, media y baja.*

Fuentes de variación	Suma de cuadrados	g.l.	Varianza estimada	F
Entre grupos	19,91	2	9,95	<0
En grupos	1.055,20	42	25,12	
Total	1.075,11	42		

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

— *Resultados de las pruebas de χ^2 para comprobar la significación estadística de las discrepancias entre las frecuencias reales y las frecuencias esperadas de ítems en cada una de las categorías de la Taxonomía de Bloom.*

Pruebas de Lengua

	χ^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: Públicos y Privados	10,98	5
Centros Urbanos de clase baja: Públicos y Privados	13,61*	5
Centros Públicos: Urbanos y Rurales	5,32	5
Centros Privados: Clase alta, media y baja	12,13	10

Pruebas de Matemáticas

	\bar{X}	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: Públicos y Privados	2,11	5
Centros Urbanos Clase baja: Públicos y Privados	2,18	5
Centros Públicos: Urbanos y Rurales	12,55*	5
Centros Privados: Clase alta, media, baja	25,52**	10

* $p < 0,05$ ** $p < 0,01$ *** $p < 0,001$ *Discusión y conclusiones*1. *Primera variable: Adecuación y relevancia de los contenidos**Pruebas de Lengua*

Tanto la media total ($\bar{X} = 22,90$) como las de los diferentes grupos están bastante alejadas de la puntuación máxima (37 puntos) si se exceptúan las correspondientes a los Centros Públicos de clase media ($\bar{X} = 27,133$). Esto significa que en general las pruebas dejan bastante que desear por lo que se refiere a la adecuación y relevancia de los contenidos. Es decir, o bien se ignoran en las pruebas contenidos relevantes o bien se presta mayor atención a los menos relevantes o ambas cosas a la vez.

Es de destacar, asimismo, la gran variabilidad de la distribución (rango = 27,5 puntos, recorrido intercuartílico = 8,74, varianza $S^2 = 35,36$). Por otro lado, al tomar en consideración las variables clasificatorias (tipos de Centros, extracción social del alumnado, enclave de los Centros) se observan grandes diferencias de variabilidad entre los diferentes grupos. Si bien a primera vista las varianzas entre Centros Privados y Públicos no son muy distintas (S^2 públicos = 32,842; y S^2 privados = 35,976), la disparidad entre urbanos y rurales es muy alta (S^2 urbanos = 17,282; y S^2 rurales = 43,597). Esto puede ser quizá debido a la dificultad de encontrar criterios adecuados que definan el estrato rural.

Algo similar ocurre en los Centros Privados cuando se clasifican en razón de la extracción social del alumnado. Se puede observar que la distribución de los Centros de clase alta es más homogénea que los de la clase media o baja. De nuevo aquí se puede poner en tela de juicio la adecuación de los criterios de definición de dichos estratos. Es muy posible que haya habido intercambios entre los estratos "media" y "baja" para Centros Privados o incluso que algunos Centros de clase alta hayan sido incluidos en los de "media".

Otras observaciones de interés se refieren a la distribución de las puntuaciones parciales respecto a los tres apartados en que se dividió el contenido lingüístico. En primer

lugar, se observa que el aspecto más descuidado en las pruebas en general es el referido al desarrollo de la competencia lingüística que los alumnos poseen como hablantes nativos. Sobre una puntuación máxima de 16, la media obtenida entre todos los colegios es de 7,38, dándose los extremos de Privados clase baja y Públicos rurales con medias de 6,06 y 5,933, respectivamente. Hay que hacer la excepción de los Públicos clase media con una $\bar{X} = 10,27$. Es, asimismo, este primer apartado el que ha registrado mayor número de puntuaciones bajas: de los 105 colegios de la muestra, 7 de ellos no han tenido siquiera en cuenta dicho apartado en sus pruebas (0 puntos) y 12 lo han considerado en grado mínimo (puntuaciones de 1 ó 2 puntos).

Todo lo anterior significa que en las pruebas de evaluación se tiende a infravalorar aquellos aspectos que conciernen al dominio por el alumno de la capacidad de usar su propia lengua. No sería aventurado pensar que tal infravaloración es reflejo de una despreocupación real por tales aspectos en la práctica docente, lo cual, por otra parte, viene a confirmar las acusaciones de intelectualismo y academicismo en que tradicionalmente ha sido objeto nuestro sistema educativo.

Los dos restantes apartados referidos al estudio reflexivo de la lengua y la aproximación a la lengua literaria, aparecen mejor tratados en las pruebas, según reflejan las puntuaciones correspondientes. En especial, la referida al último de ellos muestra gran homogeneidad existiendo escasas fluctuaciones entre las medias para los diferentes grupos de colegios con una media para toda la distribución de 8,38, muy próxima a la puntuación máxima posible para este apartado (11 puntos). Hay que tener en cuenta, no obstante, que tanto la Gramática como la Literatura son aspectos muy ricos en contenidos y, por tanto, muy fácilmente objetivables tanto en el material didáctico como en las pruebas de evaluación.

En segundo lugar, existen también diferencias notables en cuanto a la variabilidad encontrada para cada uno de dichos apartados, en perjuicio del primero de ellos. En efecto, la S^2 total para el apartado A ($S^2 = 15,89$) duplica la correspondiente al B y al C ($S^2 = 7,90$; $S^2 = 6,50$). Esta tendencia se observa también en cada uno de los grupos de Centros, destacando la gran homogeneidad de puntuaciones para el apartado C (Literatura) especialmente en los Centros Públicos de clase media ($S^2 = 0,55$) y para el apartado B en los Centros Privados de clase alta ($S^2 = 2,85$). Todo ello parece indicar que en general los profesores usan criterios discrepantes para decidir cuáles son los contenidos que deben figurar en las pruebas respecto al apartado A y, por el contrario, usar criterios similares respecto a los apartados B y C (Gramática y Literatura). Por otra parte, no sorprende este hallazgo si se tiene en cuenta la naturaleza de los contenidos de dichos apartados.

Por lo que se refiere a las distintas hipótesis de nulidad levantadas con referencia a la primera variable dependiente, los análisis estadísticos permiten las siguientes decisiones:

Hipótesis 1.^a Se rechaza la 1.^a hipótesis de nulidad tanto respecto a la puntuación total de las pruebas como para los apartados A y B de la misma. Se confirma la hipótesis de nulidad para el apartado C. Es decir, existen diferencias estadísticamente significativas entre las puntuaciones totales de los apartados A y B de las pruebas realizadas por los Colegios Nacionales y privados de Madrid. No existen diferencias significativas entre las puntuaciones de las pruebas de dichos Colegios con respecto al apartado C.

En otras palabras: las pruebas confeccionadas en los Centros Estatales tienden a ser más adecuadas en su contenido que las equivalentes de los Centros Privados que atienden niños de origen social semejante a los de aquéllos, tanto por lo que se refiere a la totalidad de la prueba como a los dos apartados A y B. No existen, sin embargo, diferencias notables de adecuación por lo que se refiere al apartado C (la Lengua Literaria).

Hipótesis 2.^a Se confirma la hipótesis nula tanto por lo que se refiere a la prueba considerada en su contenido cuanto respecto a los apartados B y C, encontrándose diferencias estadísticamente significativas respecto al apartado A. Es decir, en general no existen diferencias notables de adecuación de los contenidos entre las pruebas elaboradas en los Centros que atienden niños de clase media y los que reciben niños de clase baja (con independencia del carácter estatal o privado de los mismos si se atiende a la prueba en su conjunto). Pero si se analizan los tres apartados que la componen por separado, se encuentra que los Centros que atienden niños de clase media tienden a reflejar en las pruebas más adecuadamente los contenidos referentes al apartado A, que los Centros que atienden a chicos de clase baja, no apareciendo diferencias respecto a los dos restantes apartados, ni a la prueba en su conjunto.

Hipótesis 3.^a Se confirma la hipótesis nula. Es decir, no existe un efecto de interacción del tipo de Centro (Estatal, Privado) y la clase social del alumnado, que resulta significativo respecto a la adecuación de los contenidos de las pruebas de Lengua.

Hipótesis 4.^a Se rechaza la hipótesis nula tanto por lo que se refiere a las pruebas consideradas en su conjunto como respecto a los apartados A y C. Se confirma la hipótesis nula respecto al apartado B. Es decir, en general las pruebas elaboradas en los Centros Estatales Urbanos tienden a reflejar más adecuadamente los contenidos de la E.G.B. que las de los Centros Estatales Rurales, tanto por lo que respecta al conjunto de las pruebas como a sus apartados A y C. Sin embargo, las diferencias son prácticamente nulas respecto al apartado B (Estudio reflexivo de la Lengua).

Hipótesis 5.^a Se confirma la hipótesis nula. Es decir, no hay prácticamente diferencias respecto a la adecuación de los contenidos de las pruebas entre los Centros Privados de la Clase Alta, Media y Baja.

Pruebas de Matemáticas

Nuevamente aquí se observa que las puntuaciones de las pruebas en general se alejan bastante de la puntuación máxima posible. Tanto la media de toda la distribución como las de los diferentes subgrupos oscilan entre los 13 y 14 puntos, mientras que la puntuación máxima es de 26. De ahí que pueda afirmarse que la adecuación y relevancia de los contenidos de las mismas no es óptima ni en muchos casos ni siquiera aceptable.

La variabilidad de la distribución total es menor que la observada en las pruebas de Lengua con un rango de 18,50 y un recorrido intercuartílico de 6,70, ($S^2 = 18,02$). Asimismo, las varianzas de los distintos grupos de Centros son también más pequeñas que las observadas para la Lengua y, lo que es más significativo, no difieren notablemente entre sí.

Todo esto nos lleva a pensar que las características específicas de la materia constituyen factor originante de varianza de tal forma que la precisión de definición de los contenidos matemáticos y su clara estructuración contribuyen a que las diferencias entre las pruebas, por lo que respecta a la adecuación del contenido, sean pequeñas. Por el contrario, la mayor ambigüedad de definición de los contenidos lingüísticos, es posiblemente un factor altamente contribuyente a la gran variabilidad observada en las pruebas de Lengua, con independencia de la influencia de otros factores.

Hipótesis. Los diferentes tests estadísticos utilizados para el contraste de las hipótesis levantadas arrojan valores menores de 0 por lo que se confirman todas las hipótesis de nulidad. En otras palabras, ni el régimen de los Centros (Estatales, Privados), ni el enclave

de los mismos (Rural, Urbano), ni la extracción social de los alumnos (Alta, Media, Baja), ni sus interacciones son factores con los que pueda relacionarse la calidad de las pruebas de matemáticas en lo que se refiere a la adecuación y relevancia de su contenido.

2. Segunda variable: Peso relativo de las capacidades medidas en las pruebas

Pruebas de Lengua

En este apartado interesa analizar fundamentalmente dos aspectos:

1. Si los diferentes ítems de las pruebas se distribuyen de forma proporcional entre las seis categorías de las Taxonomías de Bloom, que han sido utilizadas como criterio, o por el contrario, predomina alguna de ellas en detrimento de las restantes.

2. Sí, cualquiera que sea dicha distribución, ésta se mantiene de forma similar entre los diferentes grupos de Centros.

En relación con el primer aspecto hay que notar el predominio de ítems referidos a la categoría de Conocimientos. Así, de un total de 1.569 ítems, 863 de ellos se refieren a esta categoría, mientras que 226 se refieren a Comprensión, 182 a Aplicación, 218 a Análisis, 1 a Síntesis y 16 a Evaluación. Por Centros, de un total de 105 Centros, 63 de ellos dedican la mitad o más de los ítems de sus pruebas a medir la categoría Conocimientos; en contraste, 15 Centros ofrecen en sus pruebas ítems que permiten ejercitar la capacidad de Evaluación y sólo uno de ellos toca la categoría Síntesis. Las categorías Comprensión, Aplicación y Análisis aparecen mejor tratadas en las pruebas, especialmente la última. Hay que tener en cuenta, sin embargo, que se han clasificado dentro de éstas todos aquellos ítems de análisis morfológico o sintáctico y es muy posible que en muchos casos la realización de los mismos no suponga un acto genuino de análisis, sino la simple aplicación mecánica de ciertas reglas y modelos previamente memorizados, con lo que la mayoría de estos ítems podrían considerarse propios de las categorías Conocimientos o Comprensión. Bajo la hipótesis de que una enseñanza bien dirigida debe cultivar por igual cada una de las seis categorías de la Taxonomía, la frecuencia teórica o esperada de ítems referidos a cada una de ellas debería ser 1/6 del total. Sin embargo, sobre esta base, las pruebas de ji cuadrado arrojan un valor de $\chi^2 = 1.909,8$, altamente significativo ($p < 0,001$) lo que deja sin confirmar aquel supuesto en nuestro caso. Agrupando las categorías de Comprensión, Aplicación, Análisis, Síntesis y Evaluación en una sola bajo la denominación genérica Capacidades Intelectuales sin hacer distinciones entre ellas, y aceptando, por tanto, la dicotomía Conocimientos-Capacidades, la frecuencia teórica debería ser de 1/2 del total de los ítems por cada categoría. Aun en este caso, las pruebas de ji cuadrado arrojan un valor de $\chi^2 = 15,71$, también significativo ($p < 0,001$).

En resumen, que el predominio de los ítems de Conocimientos en las pruebas confirman el carácter memorístico de las enseñanzas que se imparten en la mayoría de los Centros.

El segundo aspecto se relaciona directamente con el contraste de las hipótesis de nulidad levantadas para este apartado.

1.^a hipótesis. Se confirma la primera hipótesis de nulidad, dado que no se han encontrado diferencias estadísticamente significativas entre los Colegios Nacionales y Privados con respecto a la proporción de ítems referidos a cada categoría taxonómica. Las diferencias reales observadas pueden ser perfectamente atribuibles al error muestral.

2.^a hipótesis. Lo mismo puede afirmarse con respecto a esta hipótesis. Es decir, una vez

clasificados los Centros por la procedencia social de sus alumnos en las categorías media y baja, no se aprecian diferencias estadísticamente significativas entre unos y otros en la distribución de los ítems en las diferentes categorías taxonómicas.

3.^a hipótesis. Se rechaza la hipótesis de nulidad. Existen diferencias estadísticamente significativas en cuanto a la distribución de los ítems en las diferentes categorías taxonómicas entre los Centros Estatales y Privados cuando se tiene en cuenta al mismo tiempo la procedencia social de los alumnos, y viceversa, entre los Centros de clase Media y Baja cuando se tiene en cuenta al mismo tiempo su régimen jurídico (Estatad, Privado). En otras palabras, los Centros Privados de Clase Media y Baja tienden a distribuir los ítems entre las categorías taxonómicas de forma similar. Por el contrario, los Centros Públicos de Clase Media y Baja distribuyen los ítems de forma diferente, siendo mayor la proporción de ítems de Conocimientos y Comprensión y menor la de Aplicación y Análisis para los primeros que para los segundos. Por su parte, no hay diferencias significativas entre los Centros Públicos y Privados de Clase Media en cuanto a la distribución de los ítems. Sin embargo, si la hay entre los Centros Públicos y Privados de Clase Baja, siendo menor la proporción de los ítems de Conocimiento y Comprensión y mayor la de Aplicación y Análisis para los primeros que para los segundos.

4.^a hipótesis. Se confirma la hipótesis de nulidad. Es decir, las diferencias reales observadas en la distribución de los ítems en las categorías taxonómicas en los Centros Estatales Rurales y Urbanos pueden ser atribuibles al error muestral.

5.^a hipótesis. También queda confirmada esta hipótesis de nulidad. La distribución de los ítems en las diferentes categorías de la Taxonomía es similar para los Centros Privados de Clase Alta, Media y Baja, pudiendo atribuirse las diferencias reales encontradas al error muestral.

Pruebas de Matemáticas

El predominio de la categoría Conocimientos es, si cabe, todavía más acusado en Matemáticas. En efecto, el promedio de ítems encuadrados en dicha categoría es del 65 por ciento para todos los Centros (869 ítems de Conocimientos de un total de 1.337). Por grupos de Centros, solamente el de Privados de Clase Alta alcanza un porcentaje algo menor (52 por ciento) mientras que el de Públicos de Clase Media llega incluso hasta el 76 por ciento (132 ítems de Conocimientos de un total de 229). Más aún, la variabilidad es incluso menor que en Lengua, ya que de los 105 Centros, 90 de ellos (o sea, el 86 por ciento) destinan a la categoría Conocimientos la mitad o más de sus ítems. Contrastan estos altos porcentajes con la baja o nula cantidad de ítems destinados a medir otras capacidades. Así, sólo un Centro incluye en su prueba ítems de la categoría Síntesis y ninguno de ellos toca las categorías Análisis y Evaluación. A Comprensión se destinan 251 de los 1.337 (19 por ciento) y a Aplicación 216 (16 por ciento del total). Estos porcentajes apenas sufren modificación para los distintos grupos de centros formados en función de las variables clasificatorias con la única excepción de los Privados de Clase alta, donde la proporción de ítems de Comprensión es notablemente mayor que en el resto de los grupos, a expensas naturalmente de una disminución en la proporción de ítems de Conocimientos.

Bajo la hipótesis de la equivalencia de las seis categorías, la prueba de ji cuadrado arroja un valor de $\chi^2 = 2.544,4$ altamente significativo ($p < 0,001$) mayor incluso que el obtenido para Lengua. Si en cambio se acepta el supuesto de integración de las categorías 2,

3, 4, 5 y 6 de la Taxonomía en una más general (denominada Capacidades Intelectuales) y se trabaja bajo la hipótesis de la equivalencia entre Conocimientos-Capacidades, la prueba de ji cuadrado arroja un valor de $\chi^2 = 120,27$, para un grado de libertad, también significativo estadísticamente ($p < 0,001$).

Por todo ello, se puede concluir que en general los Centros de Madrid (a excepción de los Privados de Clase Alta) tienden a confeccionar pruebas de Matemáticas de 8.º en las que predominan notablemente los ítems destinados a medir la capacidad de los alumnos de memorizar información y resolver problemas de forma rutinaria, en detrimento del resto de las capacidades intelectuales tal y como son definidas en la Taxonomía. En definitiva, los datos vienen a dar testimonio una vez más de la prevalencia de lo memorístico y rutinario en las pruebas de evaluación y, por extensión, en la práctica docente actual.

Por lo que se refiere a la existencia o no de diferencias entre los distintos grupos de Centros respecto a la proporción de ítems referidos a las distintas categorías taxonómicas, el contraste de las hipótesis arroja los siguientes resultados:

1.ª hipótesis. Se confirma la hipótesis de nulidad. Es decir, no existen diferencias estadísticamente significativas entre los Centros Públicos y Privados de Madrid por lo que se refiere a la distribución de los ítems de las pruebas de Matemáticas entre las categorías taxonómicas.

2.ª hipótesis. Se rechaza la hipótesis de nulidad, o lo que es lo mismo, existen diferencias estadísticamente significativas entre los Centros respecto a la distribución de los ítems en las distintas categorías taxonómicas, diferencias atribuibles a factores asociados con la clase social (Media-Baja) en que tales Centros pueden clasificarse en razón del tipo de alumnado que atienden. En general, los Centros de Clase Media tienden a destinar mayor proporción de ítems a la categoría Conocimientos que los Centros de Clase Baja.

3.ª hipótesis. Se confirma la hipótesis de nulidad. Es decir, la consideración simultánea del régimen jurídico de los Centros (Estatad, Privado) y la clase social del alumnado (Media o Baja) no altera la distribución de los ítems en las diferentes categorías taxonómicas.

4.ª hipótesis. Se rechaza la hipótesis de nulidad. Es decir, existen diferencias estadísticamente significativas entre los Centros Públicos Urbanos y Rurales respecto a la distribución de los ítems entre las distintas categorías taxonómicas. En general, los Centros Públicos Urbanos tienden a dedicar mayor número de ítems a la categoría Conocimientos que los Centros Públicos Rurales.

5.ª hipótesis. Se rechaza la hipótesis de nulidad, o sea, que existen diferencias estadísticamente significativas entre los Centros Privados respecto a la distribución de los ítems entre las categorías taxonómicas, diferencias que son atribuibles a factores asociados con la clase social de los alumnos (Alta, Media, Baja). En general los Centros Privados de Clase Alta tienden a destinar menor cantidad de ítems a la categoría Conocimientos que los Centros Privados de Clase Media y Baja y, por el contrario, tienden a destinar mayor número de ítems a la categoría Comprensión y Aplicación. Por último, no existen diferencias estadísticamente significativas entre los Centros Privados de Clase Media y Baja respecto a tal distribución ($\chi^2 = 7,46$, g.l. = 5, $p > 0,10$).

3. Conclusiones

De todo lo precedente pueden extraerse las siguientes conclusiones:

1) En general, las pruebas de evaluación final de 8.º de E.G.B. de Lengua y Matemáticas elaboradas en los Centros no reflejan adecuadamente los contenidos de dichas materias en función de la relevancia de los mismos, en los términos en que ésta ha sido definida en la presente investigación. Tal inadecuación es más patente en las pruebas de Lengua que en las de Matemáticas.

2) Se aprecia una gran variabilidad entre las puntuaciones atribuidas a las pruebas en razón de la adecuación de los contenidos. Esto confirma la existencia de criterios dispares entre los profesores para determinar los contenidos que son relevantes y que deben figurar en las pruebas. De nuevo aquí es de notar que la mayor variabilidad corresponde a las pruebas de Lengua.

3) Por lo que se refiere en particular a Lengua, las pruebas tienden a reflejar más adecuadamente los contenidos de los apartados denominados "Estudio reflexivo de la propia competencia lingüística", y "Aproximación a la lengua literaria y sus manifestaciones históricas" (O, más comúnmente, Gramática y Literatura) que los correspondientes al apartado "Desarrollo de la competencia que los propios alumnos poseen como hablantes nativos". Asimismo, la escasa variabilidad encontrada en las puntuaciones de las pruebas con respecto a aquellos dos apartados, hace pensar que existe gran coincidencia en los criterios que los profesores usan para identificar sus contenidos más relevantes. Por el contrario, la gran variabilidad encontrada respecto al apartado citado en último lugar demuestra que no existe acuerdo entre los profesores sobre cuáles sean los contenidos más importantes dentro del mismo.

4) El análisis de las pruebas revela un predominio desproporcionado de los ítems cuya resolución exige del alumno la memorización de los contenidos o la resolución mecánica o rutinaria de los problemas, frente a aquéllos que miden las capacidades de Comprensión, de Análisis y de uso de lo ya conocido para resolver problemas nuevos. Es decir, las pruebas de evaluación final reflejan, y posiblemente fomentan, el énfasis en lo memorístico, característica secular de nuestro sistema educativo.

5) No se puede afirmar rotundamente que el régimen jurídico de los Centros (Públicos, Privados), el enclave de los mismos (Rurales, Urbanos) o la extracción social de los alumnos (Alta, Media, Baja) sean factores con los que pueda asociarse claramente la variabilidad observada, especialmente en lo que se refiere a las pruebas de Matemáticas. Sin embargo, los tests estadísticos utilizados para contrastar las diferentes hipótesis, arrojan en algunos casos resultados que, aparte de ser estadísticamente significativos, pueden contribuir al esclarecimiento en posteriores investigaciones de las causas que provocan las diferencias existentes entre las pruebas. Así, llama la atención el hecho de que las pruebas de Lengua elaboradas en Centros Públicos Urbanos reflejen en su conjunto más adecuadamente los contenidos de 8.º de E.G.B. que el resto de los Centros. De igual modo, es de resaltar que la mayor proporción de ítems de Conocimientos aparecen en las pruebas elaboradas en los Centros de Clase Media, sean Públicos o Privados, mientras que en las correspondientes a los centros Privados de Clase Alta se da la mayor proporción de ítems de Comprensión (especialmente en Matemáticas). Ello hace pensar en la hipótesis de la existencia de enfoques didácticos característicos y diferenciadores de los Centros que atienden niños procedentes de una y otra clase social. Tales divergencias pueden estar ligadas a numerosos factores: el tipo de alumno que asiste a unos y otros Centros, las diferentes exigencias de una y otra clase social, la calidad del profesorado, la organización de los Centros, las disponibilidades de medios y material de los mismos, la ratio profesor/alumno, etc., etc. y, en último término, y condicionándolos a todos, una diferente concepción de la educación. Dada la importancia de la cuestión, sería deseable emprender

nuevas investigaciones que contribuyan a confirmar o desmentir tal hipótesis y, en el primer supuesto, a identificar los factores más influyentes.

6) La mayor parte de los ítems utilizados en las pruebas de Lengua (92,70 por ciento) son de tipo ensayo. Sigue en orden de utilización el ítem de elección múltiple (3,9 por ciento). Los ítems de doble alternativa (verdadero-falso) y de completamiento de frases son solamente utilizados por los Centros Privados mientras que los ítems de elección múltiples son más utilizados por los Centros Públicos. Por otra parte, cerca de la mitad de los ítems de todo tipo utilizados en Lenguaje tienen algún error; y de los ítems de ensayo, el 47 por ciento no sirve como instrumentos de medida por los tipos de errores con que han sido redactados. Lo mismo puede decirse del 44,80 por ciento de los ítems de completar frases, del 39,60 por ciento de los ítems de correspondencia, del 19,30 por ciento de los ítems de elección múltiple y del 19 por ciento de los de doble alternativa.

En cuanto a la estructura de las pruebas, la mayoría de los Centros repiten los ejercicios de los libros de texto y confeccionan las pruebas sin otro criterio de agrupamiento de ítems que el de la materia. Ninguna prueba tiene instrucciones sobre cómo realizarla ni sobre cómo se puntuarán las respuestas.

No hay en general una estructura definida en las pruebas utilizadas en Lenguaje y son muy poco variadas en cuanto al tipo de ítem utilizado.

7) Los ítems utilizados en las pruebas de Matemáticas son fundamentalmente ejercicios de desarrollo (61,67 por ciento) y problemas tradicionales (23,09 por ciento). Los Centros Privados y, especialmente el grupo de extracción social Alta, utilizan más que los Públicos ejercicios de demostración, mientras que los problemas de tipo tradicional son utilizados por los Centros Públicos de extracción social Baja en mayor frecuencia que por cualquier otro grupo. Los Centros Privados de extracción social Alta forman el grupo que más utiliza el tipo de ejercicios de teoría y definiciones. El grupo que menos utiliza este tipo de ejercicios es el de los Centros Privados de extracción social Baja. Un tercio de los ítems aparece con errores.

En cuanto a la estructura general, las pruebas de Matemáticas muestran una simple sucesión de cuestiones o ejercicios que no responde a criterio aparente alguno.

Dirección del autor: Arturo de la Orden, Sociedad Española de Pedagogía. c/. Serrano, 127. Madrid-6.

BIBLIOGRAFIA

- BLOOM, B. S. y otros (1972): *Taxonomía de los objetivos de la educación*, tomo I, Ambito del conocimiento, (Alcoy, Marfil).
- CALONGHI, L. (1978): *Valutazione* (Brescia, La Scuola).
- EGGLESTON, J. F. y KERR, J. F. (1969): *Studies on assessment* (London, The English Universities Press).
- OLIVERA POLL, A. (1976): *La enseñanza en Madrid* (Madrid, Instituto de Estudios Madrileños).
- ORDEN, A. de la (1969). La evaluación del rendimiento educativo y la calidad de la enseñanza, *Revista de Educación*. XVIII, 206, noviembre-diciembre.
- ORDEN, A. de la (1981). Evaluación del aprendizaje y calidad de la educación, *La calidad de la educación* (Madrid, C.S.I.C.).

SUMARIO: El artículo presenta los resultados y conclusiones de un análisis de las pruebas de Lengua y Matemáticas elaboradas por los Centros de E.G.B. de Madrid para los alumnos de 8.º curso.

Se examinan tres aspectos de las pruebas: adecuación y relevancia de los contenidos, carga relativa de las capacidades medidas por las pruebas y estructura de la prueba y adecuación de los tipos de ítems utilizados.

Se contrastan las diferencias entre las pruebas elaboradas por Centros rurales y urbanos, estatales y privados y las de aquéllos enclavados en medios socioeconómicos alto, medio y bajo.

Cabe destacar entre la diversidad de resultados el predominio de los ítems tendentes a medir las conductas correspondientes a las primeras categorías (conocimiento de hechos específicos) de Bloom, y la carencia de criterios consistentes en la elaboración de las pruebas.

Descriptores: Educational evaluation, Scholastic achievement, Assessment, Educational measurement, Measurement instruments.