

Validez y fiabilidad de las pruebas de promoción

Por **ARTURO DE LA ORDEN**
Jefe de Estudios y Proyectos

ELABORACION DE LAS PRUEBAS DE PROMOCION

Una vez establecidos los objetivos, traducidos en formas de comportamiento (nociones asimiladas, formas de trabajo intelectual, hábitos mentales, operativos y sociales, destrezas, actitudes, etc.), cuya consecución por los alumnos se quiere comprobar, la primera tarea consiste en determinar las preguntas, «ítems» o elementos que han de constituir cada prueba. Las pruebas no pueden incluir toda la materia de una disciplina, ni siquiera una parte de cada unidad semanal. Es necesario elegir, seleccionar algunos aspectos o facetas de la materia entre todos los posibles. La cuestión estriba en decidir acertadamente cuáles han de ser estos aspectos.

La respuesta es clara y sencilla en términos generales: serán seleccionados como «ítems»

de las pruebas de promoción aquellas cuestiones que sean **representativas** de la asignatura en su conjunto o de la mayor cantidad posible de materia. Para determinar con cierta precisión cuáles serán estas cuestiones representativas, es absolutamente necesario estudiar y analizar a fondo cada sector de los Cuestionarios Nacionales, en orden a localizar e identificar sus puntos claves en cada curso. El conocimiento de estos puntos clave por el alumno es un índice de que domina otros muchos aspectos implícitamente comprendidos en ellos, o cuya adquisición es condición necesaria y previa.

Por otra parte, la selección de «ítems» representativos exige, a su vez, tener muy en cuenta las características de la enseñanza impartida por el maestro, que se pone de manifiesto en los programas utilizados y en los métodos seguidos.

VALIDEZ Y FIABILIDAD DE LAS PRUEBAS. ANALISIS DE LOS «ITEMS»

La calidad de una prueba de instrucción viene determinada fundamentalmente por su **validez** y su **fiabilidad**.

Se llama validez a la capacidad de una prueba para medir lo que pretende medir y no otra cosa cualquiera. Es decir, una prueba de instrucción es válida en cuanto se halla específicamente relacionada con la materia para la que ha sido elaborada y pone de manifiesto el

grado de dominio de la misma por los alumnos, en términos de los objetivos establecidos para su aprendizaje.

Se llama fiabilidad a la capacidad de una prueba para medir con exactitud y certeza, sea cual fuere el rasgo que mide, cuando se aplica en diferentes situaciones. Una prueba es fiable cuando mide con un mínimo de errores y sus medidas son relativamente constantes. Una prueba no fiable sería similar a una balanza que, al pesar en ella varias veces la misma cantidad de un determinado artículo, diera pesos cada vez diferentes.

Para determinar la calidad de una prueba, es decir, su validez y fiabilidad, pueden seguirse dos caminos:

- a) Examinar críticamente su contenido.
- b) Analizar estadísticamente los datos obtenidos de su aplicación.

En ambos casos es necesario contrastar la información que proporcionan con ciertos criterios previamente establecidos como válidos. He aquí algunos de los más utilizados para las pruebas de instrucción:

Primer criterio.—Que el contenido de la prueba, es decir, sus «ítems» o preguntas, exija de los alumnos, como respuesta, los tipos de conducta mental u operativa que el estudio de la materia trata de provocar y desarrollar. En otras palabras, el contenido de la prueba deberá traducir los objetivos de la enseñanza de cada disciplina al nivel del curso para el que se elabora.

Algunos especialistas (1) en evaluación educativa consideran este procedimiento de examen del contenido como inadecuado, y prefieren determinar la calidad de una prueba contrastando sus resultados con otras medidas o apreciaciones de la misma materia (otras pruebas, el juicio del maestro, etc.). Pero este procedimiento plantea similares problemas, ya que las mismas razones que tenemos para dudar del examen del contenido de la prueba, como criterio de calidad de la misma, son aplicables para dudar de la calidad de los criterios externos con el que vamos a contrastar los resultados.

El uso de uno u otro procedimiento dependerá de la confianza que en una situación concreta nos merezca cada uno de ellos. En muchas ocasiones, un examen crítico y detallado del contenido de la prueba, realizado por jueces competentes, puede dar mejores resultados para determinar la calidad de la prueba que cualquier otro procedimiento.

(1) Véase ROBERT L. EBEL y DORA E. DAMRIN: *Tests and examinations*, en "Encyclopedia of Educational Research", dirigida por Chester W. Harris. McMillan Co. Nueva York, 1960.

Segundo criterio: Índice de discriminación de los «ítems».—El grado en que cada una de las preguntas de la prueba es capaz de discriminar a los alumnos de diferente nivel de instrucción, es decir, poner de manifiesto lo que cada alumno sabe, separando los buenos, los regulares, los mediocres, etc., es una buena medida de la calidad de la prueba.

La mejor manera de averiguar el índice de discriminación de una pregunta es calcular el coeficiente de correlación entre la puntuación alcanzada por los alumnos en esa pregunta y la puntuación total obtenida en la prueba. Pero dada la dificultad de los cálculos que esta fórmula exige, se han ideado otros procedimientos abreviados para hallar el índice de discriminación.

Se ha demostrado que si se quiere determinar la discriminación de una pregunta, puede obtenerse un índice satisfactorio contrastando las respuestas dadas a la misma por los alumnos comprendidos en el 27 por 100 que han obtenido las mejores puntuaciones en la prueba, con las dadas por el 27 por 100 de los alumnos que han obtenido las puntuaciones más bajas.

Otra forma sencilla de averiguar la discriminación de un «ítem» consiste en hallar la diferencia entre el porcentaje de respuestas correctas dadas a la misma por el tercio superior de los alumnos y el porcentaje dado por el tercio inferior. Las preguntas que arrojan una diferencia de 40 ó superior pueden considerarse válidas desde el punto de vista de su poder discriminativo.

Tercer criterio: Índice de dificultad de los «ítems».—El índice de dificultad de una pregunta es siempre algo relativo, ya que depende del nivel de rendimiento, o mejor, de instrucción de los alumnos que constituyen el grupo a quien se formula. En general, el índice de dificultad de una pregunta se define por el porcentaje de alumnos que responden correctamente a la misma. Si se quiere incluir en una prueba «ítems» con gran poder de discriminación, la dificultad de los mismos tenderá a aproximarse a un índice 50, es decir, serán contestados correctamente por el 50 por 100 de los alumnos e incorrectamente por el otro 50 por 100. Esto implica que la puntuación media del grupo examinado alcanzará aproximadamente el valor medio de la máxima posible.

Para el análisis de los «ítems» en función de los índices de discriminación y dificultad existen tablas que facilitan extraordinariamente el proceso (2).

Cuarto criterio.—Otra forma de averiguar la calidad de una prueba consiste en determinar el grado de dispersión de las puntuaciones ob-

(2) FREDERICK B. DAVIS: *Analyse des items*. Editorial Nauwelaerts. Lovaina, 1966.

tenidas al aplicarla a un grupo de alumnos. En efecto, cuanto mayor es la variabilidad de la serie de puntuaciones, dentro de ciertos límites considerados normales, más garantías y seguridad ofrece la prueba. Se considera que una buena prueba debe dar una serie de puntuaciones con una desviación típica igual o mayor a un sexto del rango o diferencia entre la máxima puntuación posible y la mínima.

Quinto criterio.—El grado de fiabilidad de una prueba es uno de los mejores índices de su calidad, especialmente si la validez ha sido determinada de acuerdo con el primer criterio citado, es decir, a través de un examen crítico y minucioso de su contenido, por jueces competentes.

La fiabilidad de una prueba viene definida por su **constancia** o **estabilidad** (correlación entre las puntuaciones obtenidas en dos aplicaciones sucesivas de la prueba al mismo grupo de individuos), su **equivalencia** (correlación entre los resultados de la prueba y los de otra equivalente), su **consistencia interna** (correlación entre las puntuaciones obtenidas por un mismo grupo de sujetos en las dos mitades de la prueba: entre las preguntas pares y las impares, por ejemplo).

Sexto criterio.—Finalmente, es importante señalar también como índice de la calidad de una prueba la medida en que valora la velocidad en las respuestas. Las respuestas rápidas son, a veces, manifestación de un alto nivel de rendimiento; pero, en general, se considera una característica positiva de los «tests» de instrucción que los alumnos cuenten con un período de tiempo suficientemente amplio para prestar la necesaria atención a todas las preguntas.

Para determinar el tiempo óptimo concedido en una prueba se puede partir del porcentaje de alumnos que completen la prueba en un período dado o de la media del número de preguntas contestadas (bien o mal) por el grupo de sujetos.

LA RACIONALIZACION DEL PROCESO DE ELABORACION DE LAS PRUEBAS

Todas las operaciones y controles implicados en la elaboración de una buena prueba de instrucción exigen del docente una dedicación y un tiempo del que normalmente no dispone a fin de curso cuando es imperativa la realización de los exámenes promocionales.

Un buen procedimiento para hacer frente con éxito a estas exigencias consiste en confeccionar a lo largo del curso escolar un fichero de «ítems» o preguntas del cual poder extraer, al final, las más adecuadas para la prueba de promoción. Este fichero se irá enriqueciendo año tras año, y de este modo la elaboración de pruebas no constituirá problema alguno.

Para elaborar este fichero es suficiente registrar en una tarjeta cada pregunta propuesta a los alumnos en los exámenes semanales, quincenales, mensuales y trimestrales, anotando en el reverso de la misma los datos correspondientes a sus índices de dificultad, discriminación, la fecha y grupo de alumnos a quienes se aplicó y cualquier otra información que el maestro juzgue oportuno reseñar. Las fichas así cumplimentadas se archivarán clasificadas por cursos y materias.

A la hora de seleccionar las preguntas definitivas para las pruebas de progresión y promoción, se elegirán aquellas que ofrecen mayores garantías de calidad según los criterios expuestos en este artículo.

También resulta aconsejable constituir equipos entre los maestros de una localidad o comarca para la elaboración de las pruebas. Para ello, nada mejor que utilizar al máximo el marco institucional que proporcionan los centros de colaboración pedagógica. De este modo se asegura una mayor objetividad y se evitan muchos errores. Resultaría difícil que veinte o treinta profesionales, aportando cada uno su fichero, su información y su experiencia, dejen sin considerar algún aspecto importante de las pruebas.

