

CAPÍTULO 7

Diseño de pruebas para la evaluación del aprendizaje



Evaluación del aprendizaje

Instituto Internacional de Investigación de Tecnología Educativa

Diseño de pruebas para la evaluación del aprendizaje

TEMAS Y SUBTEMAS:

- 7.1 Definición del procedimiento de aplicación de pruebas
- 7.2 Tipos de pruebas para la medición aprendizaje
 - 7.2.1 Pruebas de rendimiento
 - 7.2.2 Pruebas de competencias
 - 7.2.3 Pruebas de aptitudes
 - 7.2.4 Pruebas de diagnóstico
 - 7.2.5 Pruebas intermedias de medición del aprendizaje
 - 7.2.6 Pruebas finales de medición de resultados del aprendizaje
 - 7.2.7 Pruebas para ingreso a una institución
 - 7.2.8 Pruebas para la promoción de la calidad del aprendizaje y la enseñanza
 - 7.2.9 Pruebas de egreso o titulación
- 7.3 Reglas para el diseño de pruebas para la medición del aprendizaje
 - 7.3.1 Pruebas con referencia a un criterio
 - 7.3.2 Pruebas con referencia a una norma
 - 7.3.3 Algunos de los principales errores en el diseño de pruebas de rendimiento

OBJETIVOS:

Al finalizar el capítulo, el participante será capaz de:

- ☐ Comprender el procedimiento de aplicación de pruebas.
- ☐ Clasificar los diferentes tipos de pruebas.
- ☐ Diferenciar las formas de diseño de pruebas.

CONCEPTOS CLAVE:

- | | |
|--------------------------|-------------------------------------|
| ◆ Aptitud | ◆ Prueba de diagnóstico |
| ◆ Calificación bruta | ◆ Prueba de ingreso |
| ◆ Calificación ponderada | ◆ Prueba de promoción de la calidad |
| ◆ Competencia | |

- ◆ Prueba de egreso o titulación
- ◆ Pruebas finales
- ◆ Pruebas intermedias
- ◆ Pruebas con referencia a un criterio
- ◆ Pruebas con referencia a una norma
- ◆ Rendimiento

ORGANIZADOR AVANZADO

Uno de los principales procedimientos para la medición del aprendizaje es la aplicación de pruebas, que se definen como un conjunto de preguntas o problemas relacionados con los objetivos educativos y con el contenido de la enseñanza. Existen diferentes tipos de pruebas que, de acuerdo con las características o rasgos de los individuos, se clasifican en: rendimiento, competencias y aptitudes; en relación con la programación de la enseñanza, en diagnósticas, intermedias y finales, y en relación con las políticas académicas de las instituciones, en pruebas de ingreso, de promoción de la calidad del aprendizaje y la enseñanza y finales. Las pruebas se diseñan de acuerdo con la referencia de calificación del aprendizaje, de manera que se aplican técnicas de diseño para calificar según un criterio o una norma.

7.1 DEFINICIÓN DEL PROCEDIMIENTO DE APLICACIÓN DE PRUEBAS

El procedimiento más común para la medición del aprendizaje es la aplicación de pruebas, cuyo propósito, al igual que el de otros procedimientos, es constatar que el alumno ha aprendido lo prescrito en los objetivos educativos o en los criterios de trabajo profesional o científico; o bien, que ha logrado constituir un perfil profesional, como lo propone la misión o ideario de la institución donde desarrolla su vida escolar.

Probar significa ensayar o experimentar las cualidades de una persona; examinar la exactitud o medida de las características personales de un individuo. De esta manera, se puede definir que una prueba es un instrumento para medir un conjunto de características de los individuos con propósitos determinados. Las pruebas se aplican para un sinnúmero de situaciones: de trabajo, médicas, psicológicas, etcétera, y su contenido responde directamente a los motivos de su aplicación.

En el caso de la educación institucionalizada, la aplicación de pruebas es un procedimiento enfocado a la medición del aprendizaje. Como forma de medición, las pruebas comparten las características de cualquier otro procedimiento algorítmico: se

pueden aplicar al inicio, al final o durante la enseñanza; se diseñan con referencia a una norma o a un criterio y, dependiendo de lo que se quiera medir, son útiles para examinar conocimientos, rendimientos escolares, competencias y aptitudes.

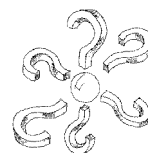
Las pruebas de medición del aprendizaje se componen de series de preguntas y problemas que el estudiante tiene que contestar y resolver, así como de reactivos a los que debe reaccionar, todo ello de manera manifiesta, de acuerdo con su capacidad y a los conocimientos que ha adquirido gracias a la educación institucionalizada o por una inclinación vocacional. La prueba brinda a los estudiantes la oportunidad de demostrar lo que han aprendido tanto de un programa de enseñanza como por iniciativa propia, y de manifestar sus creencias e ideología.

La diferencia entre pregunta, problema y reactivo, es que los dos primeros términos se relacionan con características intelectuales del individuo, es decir, con sus conocimientos y capacidad para resolver una situación problemática con fundamento en esos conocimientos; un reactivo tiene que ver con las características afectivas o los valores de los individuos, los cuales no responden a aptitudes o conocimientos de materias específicas; un reactivo describe una situación ante la cual el individuo “reacciona”, según la disposición emocional que experimenta en el momento de la prueba, o por predisposiciones basadas en las preferencias valorales personales.

Las pruebas que contienen preguntas o problemas son las más usuales en el ámbito educativo; las pruebas diseñadas con reactivos pertenecen al campo de la psicología, aunque es posible que se apliquen como complemento a los procedimientos de verificación de los perfiles de ingreso o de egreso de los alumnos.

David Ausubel (1981: 651-654), uno de los psicólogos educativos más influyentes en la teoría cognoscitiva del aprendizaje, menciona que en la larga historia del movimiento de medición en la educación, han surgido muchas objeciones a las metas de este movimiento y a los efectos producidos por técnicas particulares de medición de los resultados del aprendizaje. Algunas de estas objeciones identifican limitaciones palpables, abusos y defectos. Otras se basan en concepciones sentimentales y casi místicas del proceso educativo. Es importante explorar cuidadosamente estas objeciones y distinguir entre aquellas que se basan en limitaciones y defectos inherentes de la medición en el terreno educativo y las basadas ya sea en abusos corregibles o en capacidades que no han cristalizado.

En primer término, se argumenta que las pruebas educativas tienden a evaluar resultados de la educación, más tangibles, triviales y fáciles de medir y no los más importantes como son la comprensión genuina, la originalidad y las capacidades para resolver problemas, pensar independientemente, recuperar información, sintetizar conocimientos, etcétera. Pero esta crítica se justifica solamente en relación con las antiguas pruebas estandarizadas que medían la retención repetitiva



1. ¿Qué funciones cumple este procedimiento respecto de la evaluación del aprendizaje?

2. ¿Qué tipo de aprendizajes consideras que se pueden medir con las pruebas?

de información fáctica mediante la repetición. Debe recordarse que las pruebas objetivas existentes en la actualidad miden tanto la comprensión de principios generales como la capacidad de interpretar y aplicar conocimientos. Además, se han empleado muchas clases de instrumentos de medición para evaluar algunos de los resultados más evasivos de la educación. Entre estas pruebas figuran la observación, los informes personales, el juicio de los compañeros, las pruebas de ensayo, los exámenes orales, las muestras de trabajo, los exámenes prácticos, los artículos de investigación, etc. Es verdad que no se han ideado todavía medidas válidas de rasgos y capacidades importantes como son el estilo cognoscitivo, la creatividad, la estrategia para solucionar problemas, la flexibilidad y la sensibilidad al problema. Pero no hay razón para pensar que las dificultades que prevalecen actualmente para elaborar esos instrumentos de medición no serán superadas a la larga.

En segundo lugar, con frecuencia se argumenta que las medidas educativas no prueban el logro de objetivos *idiosincráticos* de un sistema escolar, currículo, institución o profesor determinados. No hay incompatibilidad alguna entre el uso de pruebas estandarizadas en una muestra amplia y representativa y las pruebas preparadas especialmente para un sistema escolar, escuela, currículo o salón de clases particulares. Siempre que fuese aconsejable deberían emplearse ambas clases de medidas.

En tercer lugar, las puntuaciones de pruebas y las calificaciones escolares a menudo se vuelven fines en sí mismas, y desplazan, en cuanto a importancia y supuesta validez, al conocimiento, la competencia y el aprovechamiento escolar que pretendían demostrar y representar. Cuando esto ocurre, se atrofia la pulsión cognoscitiva, los alumnos pierden interés en la materia tan pronto como son registradas sus calificaciones, y la sociedad concede mayor valor a las calificaciones o al diploma de una institución de prestigio que a los testimonios de escolaridad e idoneidad para la práctica de una profesión, intrínsecamente más válidos a largo plazo. Esta perversión de la naturaleza y función de la evaluación es en cierto sentido inevitable en una sociedad compleja, en que el significado y el propósito de los símbolos tiende a perderse con el paso del tiempo; sin embargo, la solución a este dilema no consiste en abolir las medidas de aptitud y aprovechamiento o en dejar de evaluar a las instituciones educativas, sino en aumentar el conocimiento del público acerca de la naturaleza de la medición y en combatir el culto a las puntuaciones de prueba y a los grados en el momento mismo en que aparezca.

La tendencia a considerar que las puntuaciones de las pruebas son fines de por sí y más importantes que el conocimiento al que representan es en gran parte *reflejo* de actitudes sociales indeseables sobre el valor real de la escolaridad, pero no la causa de tales actitudes ni producto inevitable de la medición y la evaluación. Así pues, el excesivo hincapié en los aspectos competitivos de las pruebas y en el empleo de puntuaciones para crear una “meritocracia” o una jerarquía pseudocientífica de personas nos dice mucho más acerca

de la clase de sociedad que sanciona tales prácticas, que sobre los abusos potenciales de la medición.

Por último, la medición y la evaluación dejan de facilitar a veces el aprendizaje o la enseñanza porque no suministran retroalimentación significativa. Esto es cierto especialmente cuando hay tan sólo exámenes finales y cuando se les informa a los estudiantes únicamente de puntuaciones, sin ningún comentario, explicación, especificación de las fuerzas y debilidades componentes, ni oportunidad de identificar o corregir los errores.

7.2 TIPOS DE PRUEBAS PARA LA MEDICIÓN DEL APRENDIZAJE

Existen diferentes tipos de pruebas para la medición del aprendizaje que se clasifican de acuerdo con las siguientes categorías: a) características o rasgos del individuo que se pretenden medir; b) función específica de la prueba, respecto de la programación de la enseñanza, y c) función específica de la prueba respecto del control escolar que ejerce la institución para el ingreso, promoción y egreso de estudiantes.

De acuerdo con las características o rasgos del individuo, las pruebas se dividen en pruebas de rendimiento, de competencias, de aptitudes y valores. Por la función que se le asigna para la programación de la enseñanza, se dividen en diagnósticas, intermedias o periódicas y finales. Para el control escolar, se dividen en pruebas de ingreso, pruebas departamentales o divisionales y pruebas de egreso o titulación.

El siguiente cuadro resume esta clasificación:

CLASIFICACIÓN DE LAS PRUEBAS PARA LA EVALUACIÓN DEL APRENDIZAJE

CATEGORÍAS	TIPOS DE PRUEBAS		
Por las características o rasgos del individuo	De rendimiento	De competencias	De aptitudes
Por las funciones que cumplen para la programación de la enseñanza	Diagnósticas	Intermedias	Finales
Por las funciones que cumplen para el control escolar	De ingreso	De promoción	De egreso o titulación

7.2.1. Pruebas de rendimiento

En primer término la palabra *rendimiento* denomina “el producto que da una cosa”; si estamos hablando de una instrucción planificada, el “producto” está predeterminado, previsto, de manera que todos los

procesos adquieren racionalidad (orden, cauce, conciencia de lo que se hace y finalidad) y están orientados por las expectativas de la forma y calidad del producto; en este sentido se habla de *eficiencia* del proceso; si de cualquier manera se obtiene un rendimiento, hablaríamos de un proceso *eficaz*, sin poner atención a los procesos que intervienen en el logro de los resultados.

En concreto, se entiende que el rendimiento hace referencia a los resultados de un proceso eficaz y eficiente de enseñanza, que no es otra cosa que el aprendizaje predeterminado por los objetivos educativos. Si se logran ambos atributos, eficacia y eficiencia, se estaría hablando de una enseñanza de buena calidad. El problema de la medición técnicamente correcta es saber, a ciencia cierta, si un proceso de enseñanza da un rendimiento predeterminado, o si éste se logra independientemente o a pesar de los procesos académicos.

De esta manera, las pruebas de rendimiento proporcionan una medida del aprendizaje; son una forma de registro cuantitativo de una muestra determinada de la adquisición de conocimientos logrado. Los resultados son interpretados a la luz de un conjunto de criterios preestablecidos o sobre la base de una norma estadística que describe las características indispensables para el aprendizaje. Esta interpretación corresponde propiamente, a una evaluación del aprendizaje. Pero, además, de acuerdo con el modelo sistémico de planificación de la enseñanza, ésta es una evaluación de resultados de la enseñanza; la información obtenida por la medición retroalimenta la eficiencia y la eficacia de la programación de la enseñanza impartida por el profesor. En este entendido, la medición del aprendizaje mediante pruebas de rendimiento escolar, es un procedimiento que provee de información para la evaluación de la enseñanza.

Generalmente, las pruebas de rendimiento miden la memorización de conceptos, la comprensión de clasificaciones, la organización de secuencias conceptuales, temporales o de situaciones históricas, fechas, etcétera, así como la comprensión y aplicación prototípica de los conceptos. Dependiendo del tipo de pregunta que se diseñe, las pruebas de rendimiento también pueden medir el aprendizaje del juicio crítico y la creatividad de los alumnos.

7.2.2. Pruebas de competencias

El concepto de las competencias humanas parte de la idea de que, en lo que respecta al desarrollo de una sociedad cultural y políticamente soberana, las personas deben ser participantes autónomas en los procesos de cambio y mejoramiento sociales, y tener la capacidad (competencia) de comprender la dinámica compleja de los procesos de desarrollo y poderla condicionar e influir en su orientación (Chinapah V. 1989: 12).

Una competencia se puede definir como la capacidad que tiene un individuo para resolver problemas y realizar actividades socialmente significativas, que determinan su calidad de vida y la de sus conciudadanos.

La diferencia básica entre la noción de rendimiento y la de competencia es que la primera está relacionada directamente con las determinaciones de los resultados de la enseñanza, es decir, con los objetivos educativos; la segunda incluye, además de lo anterior, habilidades, destrezas, conocimientos, actitudes y valores que lo capacitan para desenvolverse en un medio social determinado, con objeto de lograr su propio bienestar y el de su comunidad. Algunas de estas capacidades pueden ser resultado de la educación institucionalizada, pero otras pueden haber sido aprendidas en otros ámbitos como el familiar, cultural, grupo social al que pertenece, ambientes laborales, etcétera.

Si se acepta que el desarrollo no se limita al ámbito de la producción económica y del consumo, sino que comprende también otros aspectos como el mantenimiento de la tradición y la identidad cultural, la calidad y facilidad de las comunicaciones e interacciones entre el individuo y la colectividad, el fomento y el desarrollo de una participación ciudadana activa y crítica, y el logro y conservación de la salud, se tiene que considerar que la escala de las competencias humanas necesarias para lograr todos estos objetivos de desarrollo personal y social, aumenta correlativamente. Así como el desarrollo de una sociedad tiene múltiples facetas, las competencias y capacidades que los seres humanos deben poseer para participar activamente en las mismas, también son muy diversas y trascienden la vida escolar del individuo

Por esta razón, de entre el conjunto de competencias que puede desarrollar un individuo, sólo algunas de ellas pueden ser asumidas como responsabilidad de la educación institucionalizada. De esta manera, las pruebas de competencias tienen contenidos propiamente escolares, pero también se incluyen contenidos socialmente relevantes, como es el caso de los criterios que debe utilizar un profesional para desarrollar su práctica en contextos determinados.

7.2.3. Pruebas de aptitudes

El término aptitud define la condición o serie de características consideradas como síntomas de la capacidad de un individuo para adquirir, mediante un entrenamiento adecuado, algún conocimiento, habilidad o conjunto de respuestas (generalmente especificadas) como la capacidad para aprender un idioma o componer música. Asimismo, la aptitud refiere la capacidad del individuo para adaptarse a las condiciones del medio, es decir, el potencial que tiene, debido a su organización estructural, para enfrentarse a las condiciones generales de vida.

Esta definición sugiere dos cuestiones:

a) Que las aptitudes no son el resultado de una enseñanza o entrenamiento, sino que es una potencialidad del individuo para desarrollar un conocimiento o habilidad, que se pone de manifiesto por medio de los intercambios con el medio ambiente y las condiciones de vida. Esto significa, por ejemplo, que una persona con talento musical será capaz de componer o interpretar música si se encuentra en un medio adecuado para desarrollar esa aptitud y, en ocasiones, la desarrollará aun cuando no tenga las condiciones propicias o éstas sean mínimas; por el contrario, una persona sin ninguna aptitud musical, puede someterse a enseñanza sistematizada o entrenamientos adecuados, pero no desarrollará la capacidad de componer o interpretar música.

b) La noción de aptitud refiere una potencialidad heredada, genética o innata del individuo, es decir, depende de lo que cada quien comporta en su estructura biológica. Las aptitudes son factores de la naturaleza del individuo (biológica o espiritual) que le permiten o no desarrollar ciertas habilidades. Por ejemplo, una persona que mide 1.50 m difícilmente podrá ser un basquetbolista consumado, o una persona con dedos cortos tendrá mucha dificultad para tocar algún instrumento musical de manera virtuosa.

Existe polémica sobre la naturaleza innata de las aptitudes, pues establece un sentido determinista a las capacidades de los individuos. La idea contraria asegura que, si se encuentran las condiciones de vida adecuadas (escolar, familiar, económica, etc.), cualquier individuo puede desarrollar cualquier capacidad.

En todo caso, las aptitudes refieren un estado futuro de la persona, un “poder para aprender”, no el aprendizaje como tal. Incluyen la idea de previsión, de probabilidad de que se adquieran conocimientos y se puedan aprender ciertas habilidades.

De esta manera, las pruebas de aptitudes pretenden ser predictivas del éxito de un individuo en ciertos ambientes; en cuanto a la educación, el propósito de las pruebas de aptitudes es predecir la capacidad de los alumnos para aprender y tener éxito en formaciones profesionales específicas. Para este efecto, las pruebas se diseñan con referencia a normas relacionadas con las capacidades típicas que se requieren para poder comprender conceptos de un área de conocimientos determinada y para aprender habilidades relacionadas con estos conceptos.

7.2.4. Pruebas de diagnóstico

Son las pruebas que se aplican antes de comenzar la instrucción propiamente dicha, es decir, antes de la enseñanza de los contenidos de una materia específica. El propósito de las pruebas diagnósticas es determinar la base de conocimientos con que cuentan los alumnos, con objeto de revisar la programación de los niveles de complejidad que se han previsto para la enseñanza de la materia.

Sus funciones son:

- Determinar el nivel de conocimientos inicial de los estudiantes, previo a la enseñanza de conocimientos más complejos.
- Reprogramar los contenidos, procedimientos y materiales de enseñanza, a la luz de la verificación de los conocimientos que ya ha adquirido el alumno.
- Determinar una base de control del progreso del alumno en el transcurso de la enseñanza.
- Proporcionar una medida de diferencia entre lo que el alumno aprendió por un proceso de enseñanza específico y lo que ya ha adquirido antes de la enseñanza respectiva.

Las pruebas diagnósticas generalmente las aplica el profesor por propia iniciativa; pero también, en algunas instituciones se ha establecido como política académica la verificación de los conocimientos iniciales de los alumnos, como una base de juicio de su rendimiento efectivo y para decidir sobre la aplicación de cursos propedéuticos al inicio de la formación profesional, que solventen lagunas de conocimiento de los alumnos de primer ingreso a una institución de educación superior.

Se debe señalar que la idea de *diagnosticar* no se refiere solamente a la identificación de las carencias de conocimiento o problemas de comprensión de algún concepto específico. También está enfocada a la verificación de aquellos aprendizajes incluidos en los objetivos de la materia que ha de enseñarse y que el estudiante ya ha adquirido con anterioridad; de esta manera, no será necesario repetirlos en los contenidos ni en la programación de la enseñanza, o bien, tendrán que programarse de una forma más compleja para ahondar en lo ya aprendido hasta ese momento.

Con esta idea, las reglas generales de diseño de la prueba diagnóstica son las siguientes:

- a) Se estructura una sección en donde se cuestiona al alumno sobre los objetivos específicos de la materia, atendiendo al nivel de complejidad esperado.
- b) Se estructura una segunda sección que cuestiona los conocimientos adquiridos con anterioridad, los cuales se requieren para la comprensión de los contenidos de la materia que se ha de enseñar. Es decir, sólo se pregunta sobre los requisitos relacionados directamente con los objetivos del curso y no sobre otro tipo de conocimientos. Por ejemplo, si es necesaria la habilidad en el uso del microscopio electrónico, no se pregunta sobre historia del Renacimiento.

7.2.5. Pruebas intermedias de medición del aprendizaje

Este tipo de pruebas se aplican en el transcurso de un ciclo escolar para registrar el progreso en el aprendizaje de los alumnos, a partir de los conocimientos ya aprendidos, identificados en la prueba diagnóstica. Son pruebas que se aplican en lapsos determinados (semanalmente, mensualmente, semestralmente, etc.) y cumplen con varias funciones, entre ellas:

- Establecer una medida promedio del rendimiento de los alumnos a lo largo del curso.
- Identificar las partes de la instrucción que no han sido aprendidas o han sido mal comprendidas por los alumnos y, por lo tanto:
- Tener una base para la reprogramación de la secuencia de enseñanza, modificar los procedimientos o sustituir los materiales de instrucción.
- Definir las pautas de la repetición de secciones del contenido de la materia, la forma de retroalimentar a los alumnos, las prácticas adicionales de grupo o individuales, las tareas de estudio fuera del aula, etcétera.

Puesto que las evaluaciones intermedias tienen funciones relacionadas mayormente con la calidad de la enseñanza y, más específicamente, con la forma de programación de la instrucción, a esta evaluación también se le denomina Evaluación Formativa. El punto de medición es el rendimiento o competencia de los alumnos, pero el objeto de reformulación es la programación de la enseñanza, es decir, la selección de contenidos, la secuenciación, los procedimientos y los materiales y la presentación de las experiencias de aprendizaje.

El contenido de las pruebas intermedias corresponde a la sección o parcela del programa de la materia que ha sido enseñada en cada uno de los lapsos predeterminados, para control del progreso de los alumnos; sin embargo, es muy recomendable que cada prueba incluya reactivos de conocimientos de secciones anteriores, para mayor control y seguimiento del aprendizaje.

7.2.6. Pruebas finales de medición de resultados del aprendizaje

Las pruebas finales, como su nombre lo indica, se aplican al término del curso, cuando el profesor ha concluido con todo lo prescrito en su propia programación de enseñanza. Las funciones de la evaluación final son las siguientes:

- Evaluar el grado en que los objetivos educativos han sido logrados por los alumnos.

- Establecer una medida final de aprendizaje de los alumnos, de acuerdo con los objetivos del curso, lo que equivale a la asignación de la calificación escolar definitiva del alumno.
- Comparar los conocimientos finales con los iniciales, para determinar el progreso real en el rendimiento de los alumnos.

Las pruebas finales deben incluir reactivos de todas las secciones del programa de la materia y no sólo el conocimiento más complejo o la última parte del curso. A esta evaluación también se le llama sumativa, en la medida en que prueba en una sola ejecución la adquisición de un conjunto de aprendizajes, en el transcurso del ciclo escolar.

De acuerdo con los principios de planificación de la enseñanza, la evaluación sumativa retroalimenta a todo el proceso de programación de los contenidos de la materia y la instrucción, desde la definición de objetivos hasta la aplicación del procedimiento, los materiales de instrucción y la disposición de experiencias de aprendizaje. Recuérdese que en este paradigma de enseñanza, el rendimiento del alumno es el producto de una planificación racional, en la que a cada paso se toman las decisiones adecuadas con fundamento en la revisión de la secuencia anterior de planificación; de esta manera, los resultados del aprendizaje medidos en la prueba final, tiene como función comunicar al profesor una medida de eficiencia de su propia enseñanza y de su programación.

7.2.7. Pruebas para ingreso a una institución

Estas son las pruebas que las instituciones aplican, de acuerdo con sus políticas académicas, para otorgar el ingreso de estudiantes a los programas correspondientes de formación profesional.

Las pruebas de ingreso pueden incluir preguntas de rendimiento, de competencias y de aptitudes. Las primeras, para verificar la adquisición de los conocimientos en el nivel anterior de educación, y las segundas para establecer una medida probable de éxito en los estudios profesionales. Algunas instituciones complementan estas pruebas con otros procedimientos de evaluación como son las entrevistas a los alumnos, los avances de un proyecto personal de estudios o la verificación de algunas habilidades ya adquiridas.

7.2.8. Pruebas para la promoción de la calidad del aprendizaje y la enseñanza

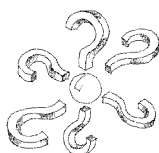
Estas pruebas se establecen, también, por política académica de la institución e instauran una verificación homogénea de conocimientos en cada nivel o área de estudios, independientemente de quién haya impartido la enseñanza de cada materia.

Las pruebas de promoción tienen como objetivo preservar un nivel de aprendizaje homologado en cuanto a la cantidad y nivel de complejidad

de los conocimientos de los estudiantes que, presumiblemente, los faculta para la comprensión de conceptos subsecuentes. Esto, a su vez, trae como consecuencia la disposición de una enseñanza también homologada respecto del conjunto de conocimientos que deben adquirir los alumnos en cada estamento del plan de estudios correspondiente, así como de los niveles de complejidad de esos conocimientos.

Esta política académica, generalmente, tiene como propósito lograr niveles de calidad estandarizados en la enseñanza y de asegurarse que los alumnos que adquieran los conocimientos mínimos necesarios para las etapas subsecuentes de su formación profesional.

El diseño de estas pruebas, en la mayoría de las instituciones que las aplican, se realiza con la participación de los profesores del área correspondiente y, en algunos casos, se han podido establecer normas de aprendizaje, como es el caso de las disciplinas tecnológicas, sobre todo las ingenierías.



1. ¿Cuáles son las relaciones que podrían establecerse entre los conceptos de rendimiento, competencias y aptitudes? Considera en tu respuesta las responsabilidades educativas de las instituciones y la capacidad de los profesores para dirigir el aprendizaje.

2. ¿Qué ventajas y desventajas observas en la aplicación de pruebas diseñadas en la institución y las que aplica un organismo social externo?

7.2.9. Pruebas de egreso o titulación

Estas pruebas están diseñadas para verificar la formación profesional de los estudiantes al final del ciclo de estudios superiores. Su aplicación, generalmente, está reglamentada por la institución y se utilizan como requisito para la certificación de conocimientos, que faculta a un estudiante egresado para el ejercicio de su profesión.

En algunas instituciones, estas pruebas son complementarias o sustitutivas de otros requisitos de egreso o titulación, que en nuestro medio son tradicionales, tales como la realización de una tesis o tesina o la integración de informes de servicio social o de la experiencia laboral.

Linn (1993) recomienda tomar precauciones al usar los resultados de las pruebas estandarizadas de rendimiento para delinear las diferencias de calidad en la educación y en la enseñanza. Menciona que en un estudio realizado en 1991 se compararon los resultados de pruebas alternativas de rendimiento en dos distritos escolares, usando instrumentos de referencia a una norma, diseñados *ex profeso*, para medir los mismos objetivos; éstas fueron homologadas con las pruebas de referencia a una norma utilizada en cada uno de los distritos, mediante un proceso de estandarización en una muestra representativa de estudiantes de distritos escolares diferentes, en los que no se acostumbraba usar pruebas estandarizadas. Se encontró que, excepto en matemáticas, los estudiantes de los distritos referidos tenían mejores rendimientos en sus propias pruebas estandarizadas que en las pruebas alternativas.

7.3 REGLAS PARA EL DISEÑO DE PRUEBAS PARA LA MEDICIÓN DEL APRENDIZAJE

Todas las pruebas tienen un patrón de elaboración con objeto de planificar el diseño de la prueba, de manera que se pueda examinar, con la mayor precisión posible, si los aprendizajes esperados han sido adquiridos por los alumnos con el nivel de complejidad especificado.

La planificación del diseño de la prueba tiene como objetivos:

✎ **Medir de manera confiable y válida el aprendizaje de los alumnos.** Esto significa que una prueba bien diseñada debe contener preguntas congruentes con lo enseñado y medir con exactitud lo ya aprendido; es decir, se debe evitar el aprendizaje en la misma prueba o la falta de respuesta por la complejidad de las preguntas.

✎ **Distribuir las preguntas** para evitar que la secuencia de disposición de los reactivos informe a los alumnos sobre las respuestas correctas de otros reactivos.

✎ **Homogeneizar el examen** para una aplicación equivalente a todos los alumnos.

✎ **Prever estados de fatiga del alumno** que le dificulten la expresión de respuestas adecuadas, entre otras cosas.

Las características del procedimiento son las siguientes:

a) *Las pruebas se aplican en un tiempo y lugar determinados.* Se espera que tanto el profesor como la institución programen de antemano la periodicidad de las pruebas, a fin de que los alumnos puedan prever sus tiempos de estudio.

b) *El cuestionamiento trata sobre las tareas que debieron aprenderse y debe ser homogéneo para todas las personas examinadas,* por lo que las pruebas se diseñan de manera que a cada uno de los alumnos se presente el mismo número y tipo de preguntas.

c) *Los alumnos están conscientes de que se someten a una situación de examen y conocen el uso que se dará a los resultados de la ejecución de tareas.*

d) *El profesor y la institución son quienes deciden el tipo de preguntas y la complejidad de las mismas, así como también las condiciones de la aplicación de la prueba y la forma de interpretación de los resultados.* En la educación moderna, la comunicación didáctica exige que el alumno conozca estas condiciones. Las pruebas elaboradas por el maestro generalmente se basan en los contenidos y objetivos propios de la materia; puede ocuparse de un tema o destreza

específicos y limitados, o de grandes segmentos de conocimientos y destrezas. Comúnmente el maestro se basa en su propio juicio y experiencia; utiliza reactivos que raramente han sido ensayados o analizados antes de pasar a formar parte de la prueba, en vez de las revisiones que hace de acuerdo con la experiencia concreta de sus clases; la prueba posee una validez muy alta, es decir, mide lo que realmente quiere medir y comúnmente no tiene como referencia más que al grupo de alumnos a los que se les aplica o a otros grupos de la escuela que ya han tomado el curso con el mismo maestro.

e) Una de las ventajas de las pruebas es que el profesor o examinador puede obtener evidencias más precisas del aprendizaje de los alumnos, ya que en la medida en que se prevén o codifican las respuestas correctas, la calificación de las pruebas dependen menos de un juicio subjetivo. En algunas instituciones se ha establecido la necesidad de asegurar la adquisición de conjuntos de conocimientos básicos, homogéneos a todos los alumnos, por lo que se han diseñado pruebas estandarizadas para lograr una calidad adecuada del aprendizaje, independientemente del profesor encargado de la enseñanza de los contenidos específicos. Se considera que este propósito se logra de una mejor manera si se aplican pruebas estandarizadas de rendimiento académico.

Para el diseño de las pruebas, se toma en consideración la base de interpretación de los resultados, es decir, el parámetro de referencia para expresar el juicio evaluativo. Esta base puede ser de dos tipos: con referencia a un criterio y con referencia a una norma.

7.3.1. Pruebas con referencia a un criterio

En este tipo de pruebas, el aprendizaje de los alumnos se mide sobre la base de sus propias habilidades para lograr un nivel determinado de comprensión, competencia o destrezas técnicas, es decir, se mide al alumno con respecto a sí mismo. Es ésta la razón por la que se recomienda efectuar una medición diagnóstica de los aprendizajes iniciales de los alumnos, previos a la impartición de la enseñanza correspondiente. Este diagnóstico permite establecer una medida inicial del aprendizaje, la cual se contrasta con la medida final, al concluir la enseñanza de una materia o conjunto de contenidos, que da como resultado el progreso real del alumno.

Las reglas de diseño de las pruebas con referencia a un criterio son las siguientes:

a) Se determinan los resultados del aprendizaje que debe medir la prueba, lo cual se encuentra en los objetivos educativos o en otros de los criterios seleccionados por la institución como parámetro del aprendizaje, como es el caso de las competencias profesionales.

b) Se determina el tipo de prueba que se aplicará —diagnóstica, intermedia o final—, ya que esto indica el tipo de conocimientos que se probarán y los niveles de complejidad.

c) *Se especifica el contenido que ha de abarcarse.* La prueba diagnóstica mide conocimientos anteriores que son prerrequisito para la comprensión de la materia que se ha de enseñar; las pruebas intermedias miden el aprendizaje de cada tema o subtema en los que se divide la materia para su enseñanza, y es recomendable incluir algunas preguntas de temas o subtemas anteriores, para verificar la asimilación u olvido de conceptos; la prueba final mide el aprendizaje de todos los objetivos de un curso. Si la responsabilidad de la medición del aprendizaje recae enteramente en el profesor, es recomendable que establezca de antemano un esbozo general de cada una de las pruebas que va a aplicar, como una forma de retroalimentación a su propia programación de la enseñanza, así como para identificar el nivel de dificultad de cada prueba, las repeticiones de preguntas que considera sustanciales para una buena comprensión y la distribución de cargas de estudio de los alumnos.

d) *Determinación del nivel adecuado de dificultad de las preguntas.* Esto depende de las prescripciones de los objetivos educativos, así como del tipo de reactivos utilizados; por ejemplo, las pruebas de ensayo son más convenientes para el nivel 6 de evaluación del dominio cognoscitivo. Debe tomarse en cuenta que no todos los comportamientos esperados como producto de los aprendizajes de los alumnos, son susceptibles de evaluarse mediante la aplicación de pruebas.

e) *Se prepara un esquema de distribución del número de preguntas por tema, según la importancia de cada uno.* Para este efecto, se recomienda hacer un borrador de prueba que abarque aquellos objetivos susceptibles de ser medidos mediante este procedimiento. En la prueba final, se debe determinar el peso específico que tendrá cada tema o subtema del programa de enseñanza, esto es, la proporción de preguntas en cada área de contenido, de acuerdo con el nivel de complejidad e importancia otorgado por el maestro a cada tema o subtema de la materia que imparte.

f) *Determinación del tipo de preguntas que se redactarán.* A este respecto, se pueden diseñar los siguientes tipos de preguntas: de ensayo, de opción múltiple, de respuesta breve, de jerarquización, de apareamiento, de columnas, de respuesta alterna. Éstos serán ampliamente descritos en la siguiente unidad.

g) *Determinación del número total de preguntas de la prueba.* Las prescripciones sobre el número de preguntas depende del tiempo que se tiene previsto para que los alumnos contesten la prueba. Para este efecto, es recomendable que se haga una prueba piloto, con el fin de establecer un número adecuado de reactivos, de acuerdo con el tiempo previsto para el examen.

Ejemplo:

El siguiente esquema de preparación de la prueba corresponde a un curso de nivel de posgrado sobre “Metodología de la investigación científica”. Este curso se ha desarrollado para apoyar a los alumnos de maestría y doctorado de una universidad, en el desarrollo de sus proyectos de investigación. Algunos de los objetivos del curso son los siguientes:

1. Los alumnos elaborarán un protocolo de investigación sobre un problema particular de conocimiento científico, que contendrá todos los elementos de planificación de la investigación empírica.

1.1. Los alumnos comprenderán el proceso de razonamiento inductivo.

1.2. Los alumnos definirán cuando menos dos problemas de investigación del área de su especialidad.

1.3. Los alumnos elaborarán hipótesis de investigación con relación a problemas de conocimiento de su área de especialidad.

1.4. Los alumnos ejemplificarán los métodos inductivos de J.S. Mill, con problemas del área de su especialidad.

1.5. Los alumnos desarrollarán procedimientos, técnicas e instrumentos de compilación de información empírica, de acuerdo con la definición de un problema de investigación.

Obsérvese que el objetivo final del curso es el diseño de una investigación, que debe expresarse en un protocolo de investigación, por lo que los objetivos 1.1 a 1.5 capacitan para una buena ejecución de la tarea esperada; de esta manera, el logro de estos objetivos de aprendizaje podrían medirse con pruebas intermedias a lo largo del curso de “Metodología...”. Obsérvese, también, que los objetivos 1.1. a 1.4. son susceptibles de medirse a través de pruebas con diferentes niveles de complejidad, pero el objetivo 1.5., por sus características, debe ser medido con otro tipo de procedimiento, ya que implica la demostración de la ejecución concreta de una tarea, del diseño de instrumentos. Asimismo, para el objetivo 1.3. se pueden diseñar dos tipos de pruebas de diferente complejidad: una de competencia para que dado un problema cualquiera del área profesional, el alumno elabore la hipótesis más plausible, y otra de rendimiento, para que el alumno identifique, deseche, mejore, etc., hipótesis ya elaboradas; la primera implicará una síntesis de conocimiento de la matriz de la disciplina correspondiente, para que el individuo sea capaz de crear la hipótesis conveniente al caso que se le presenta. Este nivel de creatividad puede rebasar la diversidad de precodificaciones (o previsión de las opciones de respuesta) que hace un profesor para el diseño de pruebas de rendimiento.

Un esquema de diseño de prueba en el que se relacionen los objetivos, el tipo de prueba, los temas y subtemas de la materia, el nivel de dificultad del aprendizaje esperado y la determinación del número de reactivos, de acuerdo con la importancia otorgada a cada tema, puede ser el siguiente:

A) OBJETIVOS	B) TIPO DE PRUEBA	C) TEMAS O SUBTEMAS RELACIONADOS	D) NIVEL DE DIFICULTAD	E) NÚMERO DE REACTIVO
1.1	Intermedia	3. La inferencia lógica <ul style="list-style-type: none"> * La lógica deductiva * La lógica inductiva * Las conexiones causales * Los problemas de la inferencia lógica 4.2. Métodos hermenéuticos 4.3. Métodos históricos	Tema 3, nivel 3 Aplicación de razonamiento inferencial. Temas 4.2 y 4.3, nivel 4 de análisis.	Tema 3 = 12 Subtema 4.2 = 6 Subtema 4.3 = 2
1.2	Intermedia	9. La definición del problema de investigación <ul style="list-style-type: none"> 9.1. La importancia de la definición del problema de investigación 9.2. Los fenómenos no explicados 9.3. La ineficacia de procedimientos 9.4. La replicación de investigaciones anteriores 9.5. La generalización de hallazgos 	Subtema 9.1, nivel 2 de comprensión. Subtemas 9.2 a 9.5, nivel 4 de análisis.	Subtema 9.1 = 4 Subtemas 9.2 a 9.5 = 12
1.3	Intermedia	10. La elaboración de hipótesis <ul style="list-style-type: none"> 10.1. El propósito de las hipótesis 10.2. Tipos de hipótesis 10.3. Formulación de la hipótesis 	Subtema 10.1, nivel 3 de aplicación. Subtema 10.2, nivel 4 de análisis. Subtema 10.3, nivel 5 de síntesis y 6 de evaluación.	Subtemas 10.1 a 10.3 = 10
1.4	Intermedia	4. Clasificaciones metodológicas <ul style="list-style-type: none"> 4.1. Los métodos inductivos <ul style="list-style-type: none"> · de concordancias · de diferencias · método conjunto de concordancias y diferencias de residuos de variación concomitante 	Nivel 5 de síntesis.	Subtema 4.1 = 10

Las pruebas con referencia a un criterio pueden ser altamente confiables si se aplican las siguientes reglas de diseño:

- Que contengan sólo lo que ha sido considerado como contenido de enseñanza o lo que se ha prescrito como materia de estudio y ha dado la oportunidad de comentarlo en clase con explicaciones complementarias, excepto lo correspondiente a las pruebas diagnósticas cuyas características permiten cuestionar conocimientos anteriores a la enseñanza de la materia correspondiente.
- Que el nivel de complejidad sea equivalente al ensayado en las clases, bajo la supervisión del maestro. Es un error muy frecuente esperar que los alumnos alcancen niveles de complejidad mayores que los enseñados, ya que los principios de aprendizaje indican que tanto lo complejo de un concepto aprendido como su generalización a otros escenarios de aplicación, sólo se logran si fueron previamente enseñados y ensayados. En otras palabras, no se puede esperar que el alumno por sí mismo pueda aplicar conocimientos complejos, si no se le ha enseñado cómo hacerlo y no se han programado las experiencias de aprendizaje para tal efecto.

7.3.2. Pruebas con referencia a una norma

Las pruebas con referencia a una norma son aquellas en las que el aprendizaje de un alumno se juzga en comparación con los aprendizajes estandarizados de su grupo de compañeros; el estándar se define como un promedio de ejecución, el cual, a su vez, puede dividirse en diferentes niveles de complejidad.

Estas pruebas se basan en un contenido y objetivos comunes a varios grupos de alumnos a los que se les enseña una misma materia, o que provienen del mismo nivel escolar o de escuelas afines; se ocupan de grandes segmentos de conocimientos o de destrezas, y comúnmente las preguntas se agrupan de acuerdo con una destreza o un tema particulares; son elaboradas por redactores, revisores y preparadores profesionales; incluyen preguntas que han sido ensayadas, analizadas y revisadas antes de pasar a formar parte de la prueba; si está bien elaborada, la prueba posee un alto grado de confiabilidad, es decir, si se aplica de manera repetida al mismo sujeto, éste tiende a obtener la misma calificación, si es que no se ha sometido a instrucción adicional; proporciona normas para diversos grupos que, en términos muy generales, son representativas de la ejecución de la muestra de población a la que pertenece el individuo que se somete a la prueba.

Las pruebas estandarizadas están dirigidas a medir los conocimientos que deben ser aprendidos por el común de una población afín; éstas tienden a tener una cualidad predictiva (aunque éste no sea su propósito principal) pues permiten correlacionar la calificación obtenida por un sujeto con el comportamiento futuro del mismo individuo. Pueden contener los conocimientos indispensables para

un profesional, o los conocimientos que no deben ser ignorados por alumnos de materias iguales o equivalentes.

Puesto que su propósito es medir los conocimientos de poblaciones grandes, estas pruebas tienden a detectar más las aptitudes o competencias que los conocimientos específicos, dada la variedad de enseñanzas que puede recibir cada individuo.

Las pruebas estandarizadas poseen tres características en común:

- Tienen una serie fija de conceptos;
- Tienen una serie de instrucciones específicas para su administración y puntuación, que posibilita la aplicación homogénea de la prueba, y han sido probadas en muestras representativas de un universo determinado de individuos, de las cuales se obtienen las normas de referencia.

Los conceptos y procedimientos uniformes de administración y calificación hacen que estas pruebas sean particularmente útiles para comparar las puntuaciones individuales o grupos de puntuaciones obtenidas en diferentes tiempos y lugares; en tanto que los datos resumidos de administraciones anteriores proporcionan los niveles o normas contra los cuales se pueden hacer comparaciones.

Es importante entender que, comúnmente, el objetivo de las pruebas estandarizadas *no* es medir el progreso mediante un curso particular de instrucción; su finalidad principal no es medir el logro de determinados objetivos educativos; sin embargo, tienen, de hecho, funciones evaluadoras, que se centran en la idea de que el aprendizaje logrado por un alumno es una información relevante acerca de la calidad de la enseñanza que proporciona un profesor o una institución. Esto es muy común en relación a la imagen que proyecta el profesor o institución correspondientes, cuyos alumnos logran las mejores calificaciones en exámenes donde se aplica una norma para asignar dicha calificación.

Para su diseño, además de lo señalado para las pruebas con referencia a un criterio, las pruebas estandarizadas deben someterse a un proceso de normalización estadística; es decir, a la obtención de las normas de calificación asignadas a las ejecuciones de los individuos que se someten a la prueba correspondiente. El procedimiento de normalización es el siguiente:

a) Se prepara una primera versión de la prueba y se aplica a grupos semejantes a aquellos para los cuales se ha elaborado. En este primer ensayo se examinan cuestiones como la redacción de las preguntas, el nivel de complejidad, la claridad y precisión de las instrucciones de respuesta y tiempo de aplicación, entre otras cosas. Es decir, se hace una primera prueba de validez del contenido de la prueba.

b) Se hace un análisis de las preguntas de las instrucciones, del tiempo de aplicación, etc., a la luz de los resultados de la aplicación de ensayo; esta revisión se realiza para prever las medidas de confiabilidad de la

prueba, para homogeneizar la aplicación y para que el procedimiento no influya sobre los resultados de la medición.

c) Se hace una versión revisada de la prueba.

d) La versión final se aplica a una muestra representativa del grupo de población para el que se ha diseñado la prueba. En este paso se pueden realizar las pruebas de validación de criterio y de constructo hipotético, así como la obtención de los índices de confiabilidad mediante la aplicación repetida de la prueba a poblaciones semejantes o mediante la división de la misma prueba, para correlacionar las puntuaciones en las dos mitades.

e) Se califican las pruebas, se tabulan las puntuaciones y se obtienen las normas estadísticas de respuesta, lo cual proporciona un parámetro para la interpretación de las puntuaciones obtenidas por un individuo, cuando se somete a una prueba estandarizada. La estandarización determina los niveles de dificultad de las preguntas, a las cuales se les asigna una ponderación de acuerdo con esta dificultad, es decir, las respuestas correctas con algunas preguntas pueden tener una mayor puntuación que otras.

Las respuestas de un individuo a una prueba estandarizada determinan su *calificación bruta*. Esta calificación carece de significado si no se compara con el parámetro estandarizado, o sea con la ejecución promedio del grupo de individuos del que se obtuvo la puntuación estándar. Para que la calificación del sujeto sea válida y confiable, debe tener características afines al grupo del que se obtuvo el estándar. La comparación permite determinar la posición exacta del individuo con respecto a este grupo. Para este efecto, las calificaciones brutas son transformadas, mediante diferentes procedimientos estadísticos, en *calificaciones ponderadas*, cuyo propósito es comparar las calificaciones obtenidas en diferentes pruebas, expresándolas en un mismo tipo de unidad. Existen tres tipos de calificaciones ponderadas: de edad o cronológica, de percentiles y tipificada; la primera es común para las pruebas psicológicas de desarrollo del individuo, las dos últimas son las que se aplican en la medición del aprendizaje.

f) Se hace una segunda aplicación de ensayo para establecer el grado de confiabilidad de las normas y las correlaciones con otras medidas (por ejemplo, entre el rendimiento y las aptitudes).

g) Se preparan manuales para administrar de una manera homogénea la prueba y para interpretar las puntuaciones de cada individuo que se somete a ella.

7.3.3. Algunos de los principales errores en el diseño de pruebas de rendimiento

Las reglas anteriores de diseño de pruebas se han emitido para evitar una serie de errores que son muy comunes cuando se diseñan pruebas para la medición del aprendizaje. Los errores más frecuentes son los siguientes:

a) Las preguntas se disponen de tal manera que los alumnos experimentan un aprendizaje a través de las respuestas sucesivas, por lo que se les facilita contestar los reactivos más complejos o los relacionados con las preguntas aprendidas.

b) Las instrucciones están redactadas de tal manera que los alumnos no entienden cuál es la respuesta que se les pide. La sintaxis de las preguntas es confusa, no se dan los datos completos, incluyen intenciones paradójicas o capciosas, etcétera.

c) Las preguntas sólo evocan la memorización de los conceptos, lo cual le resta validez y confiabilidad a la prueba respecto a los objetivos educativos más complejos.

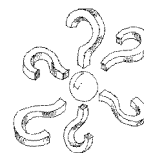
d) La prueba es de tal extensión que provoca fatiga en los alumnos y les dificulta la expresión de las respuestas adecuadas. Es frecuente que algunos profesores sometan a sus alumnos a pruebas de 4 ó 5 horas, cuando en el medio laboral, la solución de los problemas profesionales se realiza con otros parámetros, como la consulta a los colegas o la secuenciación y prueba continua de las soluciones.

e) El nivel de complejidad de las preguntas es diferente a lo especificado en los objetivos y a lo enseñado por el profesor.

f) Los alumnos desconocen el procedimiento de calificación de la prueba.

Desde que se inició la historia de las pruebas estandarizadas, éstas han sido vistas como un procedimiento eficaz de predicción de los rendimientos futuros de los individuos. De hecho, en Estados Unidos, las calificaciones altas en algunas de estas pruebas son consideradas como un requisito para la aceptación de alumnos en programas de licenciatura y posgrado. Esto obedece a políticas académicas de preservación de una calidad lograda por los programas respectivos, que incluye la aceptación de los mejores estudiantes potenciales. En este contexto, las pruebas de rendimiento se utilizan para asegurarse de que los alumnos han adquirido los conocimientos necesarios para estudios más complejos, pero también para tener un índice de predicción del éxito en los estudios; esto permite a las instituciones ahorrar recursos, conservar índices altos de egreso y graduación, asegurar que los estudiantes admitidos conserven calificaciones altas en el transcurso de su vida escolar, etcétera, todos los cuales son criterios de calidad. Asimismo, se aplican pruebas que miden otro tipo de rasgos de los estudiantes, correlacionadas con rendimiento, para obtener confiabilidad en los criterios de selección de los alumnos.

Paul Ramsden (1986) menciona que los procedimientos que se usan para medir los estándares de calidad han llegado a ser más importantes que las cosas que miden, de manera que se ha perdido de vista el propósito de la medición. En su opinión, la transformación de los medios en fines, en la medición del aprendizaje de los alumnos, ha desvirtuado los estándares de calidad. En particular, porque los medios



1. ¿En qué ocasiones utilizarías pruebas de rendimiento estandarizadas, en lugar de pruebas referidas a un criterio?

2. ¿Cuál sería la utilidad de estandarizar las pruebas de la formación profesional?

3. ¿Cómo se complementarían con las pruebas realizadas por el profesor o la institución?

para la medición son cuantitativos y se ha olvidado la naturaleza cualitativa de los cambios educativos —el hecho de que el aprendizaje en la educación superior invariablemente incluye aprehender el significado del conocimiento además de acrecentarlo—.

Ramsden describe los resultados de un estudio realizado en Estados Unidos, avalado por el *Council for National Academic Awards*, que se han interpretado como evidencia de que las instituciones del sector público son más eficientes para producir egresados que las universidades. Esta interpretación está basada en el hecho de que los politécnicos y los *Colleges* que reciben a los estudiantes con menores calificaciones, producen proporcionalmente, más graduados con honores de primera clase que las universidades, a pesar de que la enseñanza de las universidades produce una mayor cantidad de graduados de excelencia.

Argumenta que se pueden tener dudas sobre esta clase de análisis cuando se observan la falta de confiabilidad y la posible invalidez de las variables utilizadas como indicadores de calidad, es decir, se ha comprobado que las altas calificaciones no predicen con exactitud el rendimiento futuro de los alumnos, además de que la percepción del valor de los certificados obtenidos en diferentes instituciones, es muy relativo, si se toman en cuenta las condiciones de la opinión pública.

A su juicio, la dificultad real de comparar la calidad poniendo atención a la calificación de los resultados del aprendizaje no es un problema técnico del procedimiento de medición, sino que está relacionada con el hecho de que se cree que los indicadores cuantitativos pueden mostrar objetivamente un rendimiento, cuando esto es en realidad un concepto de valor. Esto incluye aceptar la incertidumbre y la falibilidad del juicio humano al decidir los estándares de calidad del aprendizaje; significa, también, que los profesores deben tomar mayor responsabilidad profesional en la indagación de las mejores formas de aprendizaje de sus alumnos y aplicar sus hallazgos en la enseñanza.

RESUMEN

La aplicación de las pruebas de rendimiento es el procedimiento más usual para la evaluación del aprendizaje. Las pruebas pueden clasificarse, de acuerdo con las características o rasgos de los individuos, en: rendimiento, competencias y aptitudes; según periodo de instrucción en el que se aplican, en: pruebas diagnósticas, intermedias y finales, y en relación con las políticas académicas de las instituciones, en pruebas de ingreso, de promoción de la calidad del aprendizaje y la enseñanza y en pruebas finales. Las pruebas diagnósticas tienen como propósito establecer los conocimientos ya adquiridos por el estudiante

que son relevantes para el contenido de la materia de estudio específico; las intermedias se aplican cada determinado tiempo al concluir algunas de las secciones en las que se ha programado la impartición de la materia; la prueba final se aplica una vez concluida la instrucción de la materia y tiene como propósito, evaluar el rendimiento del alumno respecto al contenido y complejidad de los objetivos educativos prescritos para esa materia específica. Las pruebas se diseñan de acuerdo con la forma de calificación de las mismas: con referencia al criterio de rendimiento prescrito en los objetivos o con referencia a una norma estandarizada de calificación del grupo afín al estudiante. Las pruebas deben cumplir con dos criterios de diseño: la validez, que indica si la prueba realmente mide lo que pretende medir, y la confiabilidad, que proporciona un índice de precisión de la medición. El diseño de pruebas tiene un conjunto de reglas específicas que evitan el aprendizaje en la propia prueba, la fatiga del estudiante, la facilitación de las respuestas por información contenida en la misma prueba y la variabilidad en la complejidad de las respuestas respecto a lo enseñado por el maestro.



AUTOEVALUACIÓN

1. ¿Por qué no es posible medir con una prueba el aprendizaje del siguiente objetivo?

“A. Los alumnos elaborarán un protocolo de investigación sobre un problema particular de conocimiento científico, que contendrá todos los elementos de planificación de la investigación empírica”.

2. Escribe en el paréntesis a la derecha de los siguientes enunciados, una (C) si lo descrito corresponde a pruebas con referencia a un criterio y una (N) si son características de pruebas con referencia a una norma:

El aprendizaje de un alumno se juzga en comparación con los aprendizajes estandarizados de su grupo de compañeros. ()

Están dirigidas a medir conocimientos que deben ser aprendidos por el común de una población afín. ()

El aprendizaje de los alumnos se mide sobre la base de sus propias habilidades. ()

Son probadas en muestras representativas de un universo determinado de individuos, de las cuales se obtienen las normas de referencia. ()

Determinan los resultados del aprendizaje que debe medir la prueba, lo cual se encuentra en los objetivos educativos. ()

Determinan el nivel adecuado de dificultad de las preguntas equivalente al ensayado en las clases. ()

3. Escribe en el paréntesis a la derecha de los siguientes enunciados, una (D), si las siguientes características corresponden a una prueba diagnóstica; una (I) si corresponden a una prueba intermedia, y una (F) a una prueba final.

Mide el aprendizaje de los contenidos de todos los objetivos estipulados para la enseñanza de una materia. ()

Mide el aprendizaje de conocimientos previos a la enseñanza de una materia específica. ()

Mide el aprendizaje de contenidos que capacitan a los alumnos para el aprendizaje de los conocimientos más complejos de una materia específica. ()

4. ¿Qué diferencias existen entre las pruebas para la promoción de la calidad de la enseñanza y el aprendizaje y las que elabora un maestro?

RESPUESTAS

1. La naturaleza de la tarea que se pide al alumno implica un componente de creatividad que hace muy amplio el tipo de respuestas correctas posible; por lo tanto, una prueba con preguntas no sería ni válida ni confiable.

2. El aprendizaje de un alumno se juzga en comparación con los aprendizajes estandarizados de su grupo de compañeros. (N)

Están dirigidas a medir conocimientos que deben ser aprendidos por el común de una población afín. (N)

El aprendizaje de los alumnos se mide sobre la base de sus propias habilidades. (C)

Son probadas en muestras representativas de un universo determinado de individuos, de las cuales se obtienen las normas de referencia. (N)

Determinan los resultados del aprendizaje que debe medir la prueba, lo cual se encuentra en los objetivos educativos. (C)

Determinan el nivel adecuado de dificultad de las preguntas, equivalente al ensayo en las clases. (C)

3. Mide el aprendizaje de los contenidos de todos los objetivos estipulados para la enseñanza de una materia. (F)

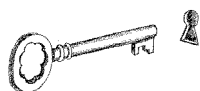
Mide el aprendizaje de conocimientos previos a la enseñanza de una materia específica. (D)

Mide el aprendizaje de contenidos que capacitan a los alumnos para el aprendizaje de los conocimientos más complejos de una materia específica. (I)

4. Cualquiera de las siguientes respuestas u otras afines son correctas:

Generalmente el maestro sólo mide el aprendizaje de su materia, con su propio método didáctico; las pruebas generales que se aplican a todos los estudiantes aseguran una calidad homogénea o mínima, común a todos.

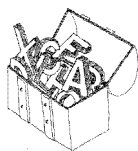
Las pruebas que estandariza una institución pueden hacer referencia más general a las competencias profesionales que se espera de los estudiantes; en cambio, las pruebas de los profesores, aunque pueden medir destrezas complejas, pueden parcializar el conocimiento.



LECTURAS COMPLEMENTARIAS

Grohund, N.E., "Cómo planificar el test", en *Elaboración de test de aprovechamiento*, Trillas, México, 1997, pp. 27-40.

Smet V.N.M., Valenzuela M.N. Lugo C.M.G., *Una nueva tecnología para la evaluación del desempeño académico a nivel bachillerato: el sistema de evaluación del aprendizaje (SEA)*, ponencia, 1999.



GLOSARIO

Aptitud: Condición o serie de características consideradas como síntomas de la capacidad de un individuo para adquirir, mediante un entrenamiento adecuado, algún conocimiento, habilidad o conjunto

de respuestas. Asimismo, refiere la capacidad del individuo para adaptarse a las condiciones del medio, es decir, la capacidad que tiene, debido a su organización estructural, para enfrentarse a las condiciones generales de vida.

Calificación bruta: Medición de las respuestas de un individuo a una prueba estandarizada.

Calificación ponderada: Medición de las respuestas de un individuo a una prueba estandarizada comparada con las medidas estandarizadas.

Competencia: Capacidad de comprender la dinámica compleja de los procesos de desarrollo y poderla condicionar e influir en su orientación.

Prueba: Instrumento que se aplica para la medición de un conjunto de características de los individuos con propósitos determinados.

Prueba de diagnóstico: Pruebas que se hacen antes de comenzar la instrucción propiamente dicha. Su propósito es determinar la base de conocimientos con que cuentan los alumnos.

Pruebas de ingreso: Pruebas que las instituciones aplican, de acuerdo con sus políticas académicas, para otorgar el ingreso de estudiantes a los programas correspondientes de formación profesinal.

Pruebas de promoción de la calidad del aprendizaje y la enseñanza: Establecen una verificación homogénea de conocimientos en cada nivel o área de estudios, independientemente de quien ha impartido la enseñanza de cada materia.

Pruebas de egreso o titulación: Pruebas diseñadas para verificar la formación profesional de los estudiantes al final del ciclo de estudios superiores

Pruebas finales: Pruebas que se aplican al término del curso para evaluar el grado de aprendizaje de lo contenido en los objetivos educativos; establecen una medida final de rendimiento de los alumnos.

Pruebas intermedias: Pruebas de rendimiento que se aplican en el transcurso de un ciclo escolar para registrar el progreso en el aprendizaje de los alumnos, a partir de los conocimientos ya aprendidos, identificados en la prueba diagnóstica.

Pruebas con referencia a un criterio: Pruebas diseñadas para medir el aprendizaje de los alumnos sobre la base de sus propias habilidades para lograr un nivel determinado de comprensión, competencia o destrezas técnicas, es decir, se mide al alumno con respecto a sí mismo.

Pruebas con referencia a una norma: Pruebas que han sido probadas en muestras representativas del universo de población para establecer una norma de ejecución de tareas y un criterio de calificación del individuo, respecto a dicha norma.

Rendimiento: Resultado de un proceso de enseñanza, que hace referencia al aprendizaje predeterminado por los objetivos educativos.



OTRAS FUENTES DE CONOCIMIENTO

Puedes encontrar ejemplos diversos de utilización de pruebas de rendimiento en las siguientes direcciones electrónicas:

Una introducción a los términos y conceptos asociados con la medición estandarizada aplicada a través de medios electrónicos (*Computer based assessment*):

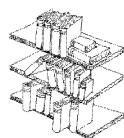
<http://www.derby.ac.uk/schools/business/bmc/cbal.html>

Asimismo, existe en la red sociedades de consultoría para el diseño de pruebas de referencia a un criterio como: *Criterion Referenced Consulting Associates International*

<http://www.nis.za/crcn/page7.htm>

El *homepage* para consultoría en la instrucción con referencia a criterios (*Criterion Referenced Instruction CRI*)

<http://www.nis.za/crcn>



BIBLIOGRAFÍA

- AUSUBEL, David, *Psicología educativa. Un punto de vista cognoscitivo*, Trillas, México, 1981.
- CHINAPAH, Vinayagum; Jan-Ingvar Löfstedt y Hans Weiler, “Desarrollo integrado de las competencias humanas y planeamiento de la educación”, en *Perspectivas*, vol. XIX, no. 1, 1989, 11- 33.
- COSTA Neiva, Kathia María, *Manual de pruebas de inteligencia y aptitudes*, Plaza y Valdés, UIA, México, 1996.
- DICCIONARIO de Psicología, Fondo de Cultura Económica, México, 1982.
- GRONLUND, Norman E., *Elaboración de tests de aprovechamiento*, Trillas, México, 1980.
- LINN, Robert L., “Educational assessment expanded expectations and challenges”, *Educational Evaluation and Policy Analysis*, 15(1):1-16.

- RAMSDEN, Paul, "Students and quality", en Moodie, Graeme C. (ed.), *Standards and criteria in Higher Education*. Gran Bretaña, SRHE& NFER-Nelson, 1986.
- THORNDIKE, Robert L. y Elizabeth Hagen, *Tests y técnicas de medición en Psicología y Educación*, Trillas, México, 1975.