



OE
2

SERVIÇO PÚBLICO FEDERAL

DISTRIBUIÇÃO

Una nuevo enfoque de la evaluación
del Aprendizaje
Universidad de Chile

SERVIÇO PÚBLICO FEDERAL

UNIVERSIDAD DE CHILE

Instituto de Investigaciones Estadísticas

por Sra. Lucy Bianchi

Hablar de un nuevo enfoque de la evaluación educacional implica, primordialmente, establecer en forma clara y precisa, el significado de los términos "evaluación" y "medición" educacionales, ya que, en el presente, la evaluación del aprendizaje está fundamentada en la medición. Sólo en esta forma constituye un proceso científico, mediante el cual se trata de alcanzar el máximo de objetividad en sus resultados.

Basándonos en el concepto general de medición de Campbell, quien dice que ésta consiste en "asignar números a los sujetos o sucesos, de acuerdo a reglas", entendemos por medición educacional "la representación cuantificada del grado en que un alumno posee determinado rasgo o cualidad".

Las pruebas y exámenes son el instrumento de medición educacional, ya que sus resultados describen, mediante una cifra, el grado en que los estudiantes poseen ciertos conceptos, habilidades y destrezas. La medición educacional científica tiende a centrarse en un rasgo o característica estrecha, específica y bien definida y luego se esfuerza por determinar con precisión, en qué medida un alumno posee ese rasgo. Si nuestra medición se ha efectuado cuidadosamente, con instrumentos adecuados, los resultados numéricos no cambian. Por lo tanto, dentro de la medición educacional es muy importante el instrumento de medida.

Este debe ser de tal calidad, que mida a todos los individuos con la misma vara y, para que sea adecuado, debe poseer determinadas propiedades: VALIDEZ y CONFIABILIDAD.

La evaluación educacional, por otra parte, constituye un proceso continuo en que el profesor, utilizando información derivada de diversas fuentes, emite un juicio de valor. Esta información puede obtenerse usando tanto instrumentos de medición, como otras técnicas que no necesariamente den resultados cuantitativos; estas técnicas, sin embargo, deberán ser lo más objetivas posible para proporcionar una base sólida a la evaluación.

Si la medición educacional y las pruebas y exámenes como sus instrumentos, de acuerdo a los conceptos anteriores, constituyen un aspecto fundamental dentro de la evaluación del aprendizaje, podemos preguntarnos:

¿Es posible describir en forma numérica los atributos humanos?

¿Es necesaria la medición del aprendizaje?

Thorndike ha dicho que los atributos humanos pueden medirse porque las características individuales se dan en diversos grados y todo lo que existe en alguna cantidad es susceptible de ser medido. Sin embargo, decir que todo lo que existe puede medirse no significa que sepamos de antemano cómo lo mediremos, que esta medición pueda hacerse en forma fácil y rápida o que todo pueda ser medido mediante pruebas. El problema de medir las diferencias entre las características individuales es esencialmente el de desarrollar medios que ayuden a nuestros sentidos, que refinen lo que simplemente hemos percibido, que hagan más exactos nuestros informes cuantitativos, más fáciles de reproducir, más confiables. Todo esto corresponde al constructor de pruebas educacionales y a los docentes que construyen pruebas para evaluar a sus alumnos.

La pregunta: ¿son necesarias las pruebas?, podría ser respondida afirmativamente aún por los que las critican y consideran que debería prescindirse de ellas, ya que difícilmente encontramos a un profesor a quien no importe la calidad de la enseñanza que está impartiendo, los conocimientos de sus alumnos, su preparación previa, su propio éxito como profesor y la efectividad del proceso total del aprendizaje. Como la calidad es materia de grado, necesitamos una forma de medirla; si abandonamos las pruebas, necesitaremos un medio mejor que éstas. En la actualidad, la mayor parte de las críticas van dirigidas a mejorar los instrumentos de medida, no a su supresión.

De lo dicho deriva la importancia que en todos los niveles educacionales se confiere actualmente a la construcción de las pruebas y exámenes; éstos, al ser considerados instrumentos científicos de medición, deben cuantificar las características en estudio, lo que, a su vez, implica obtener ventaja de los beneficios que las operaciones con números y el razonamiento matemático permiten.

La preparación de toda prueba incluye cierto número de decisiones. Por lo tanto, debe dedicarse tiempo y trabajo a la construcción de los exámenes, para así poder estudiar cuáles decisiones se tomarán y para reflexionar acerca de las causas que influirán en ellas.

Se compara, en ciertos aspectos, la planificación necesaria para construir una prueba, a las etapas de una investigación. Constituye un problema científico. Toda investigación se realiza porque existe un problema. ¿Cuál será el problema en nuestro caso?

Propósito de la Prueba: ¿Qué medir y para qué medir? ¿Mediremos conocimientos, aptitudes, intereses u otros aspectos de la personalidad? ¿Con qué objeto efectuaremos esta medición? ¿Para efectuar una evaluación de los conocimientos, para predecir el comportamiento futuro, para evaluar un método, para seleccionar los mejores alumnos, para evaluar al profesor?

Una vez precisado el propósito del examen, el constructor del examen se preguntará:

¿A quiénes mediremos? ¿A los alumnos del grupo práctico de los días Lunes? ¿Al I, II o III año de Química en la Escuela de Agronomía? ¿A los que deben rendir examen de repetición? ¿A un grupo de postulantes a primer año?

Al responder la pregunta "¿a quiénes mediremos?" estamos fijando la "población" o universo al cual administraremos el examen.

Cada uno de estos grupos necesitará pruebas diferentes, ya que constituyen en sí una población de sujetos distinta una de otros.

En toda investigación existe un investigador u observador. En la construcción de un examen considerado científicamente, éste será la persona que efectúa la medición. En la mayoría de los casos, este rol corresponde al profesor del curso.

Las ventajas educacionales que resultan del hecho de que el profesor construya su prueba, son las siguientes:

- Le ayuda a clarificar y definir los objetivos educacionales de su curso.
- Este tipo de exámenes estará más relacionado con el contenido específico y con los objetivos de un determinado curso, que un examen preparado por otras personas.
- Al estar la construcción de pruebas y la enseñanza en manos de una misma persona, se produce una mayor integración del proceso educacional total.

Esto no significa que todas las preguntas de un examen sean producto original del profesor. Puede y debe existir un intercambio, tanto de esquemas de pruebas (tablas de especificaciones), como de preguntas, entre profesores de cursos similares. En esta forma, no sólo se reduce el trabajo del constructor, si no que también es posible utilizar preguntas cuyas características estadísticas se conocen de antemano. Las mejores preguntas pueden y deben usarse más de una vez, siempre que se mantengan las precauciones de seguridad apropiadas para que ellas no sean copiadas o memorizadas, pasándose de un grupo de alumnos a otro.

Además de intercambiar preguntas, los profesores pueden formar su propio "cardex" de items, eligiendo los más apropiados en el momento de construir su prueba. Estas preguntas tendrán gran valor si se conoce su grado de dificultad y de discriminación. Existen, también, pruebas que deberán administrarse a grandes grupos de alumnos, como sucede en los exámenes de ingreso a la universidad. Actualmente, estos exámenes son preparados por comités que incluyen a expertos en evaluación educacional y a profesores especialistas en la materia.

Muy importante en la planificación de todo examen es el análisis previo de los tópicos que se incluirán en él. Pero, como el rendimiento educacional consiste, no sólo en la adquisición de determinados conocimientos, sino que también en el desarrollo de habilidades y destrezas para desempeñar tareas específicas, el estudio del contenido de una prueba deberá también abarcar los "objetivos educacionales" representados por estas habilidades o destrezas; esto implica que el profesor, al evaluar, debe tener claro qué desea medir, definirlo en acciones, es decir, las variables incluidas deben ser expresadas en términos adecuados, susceptibles de ser comprobadas en conductas observables.

Todos los aspectos mencionados, más el tipo de preguntas que se utilizará, su número y su nivel de dificultad, son aspectos que la evaluación actual del aprendizaje considera necesario analizar y precisar antes de construir el instrumento de medida. La planificación de éste concluye con la elaboración de una tabla de especificaciones, que puede considerarse como un esquema de la prueba en el cual se encuentran las áreas de contenido en una dimensión y las destrezas y habilidades intelectuales en la otra.

Construcción de las preguntas

Construir preguntas constituye un arte. Requiere una combinación de habilidades específicas que exigen, para su desarrollo, de práctica extensa y revisión crítica continua.

Es una labor esencialmente creativa. Cada pregunta representa nuevos problemas y nuevas oportunidades.

Los que no han escrito preguntas que traten de ajustarse a determinados standards en cuanto a calidad, difícilmente apreciarán la dificultad de su construcción. Contrasta esta posición con el criterio generalizado de que es posible construir una prueba en una o dos horas.

El escritor de preguntas deberá:

Dominar la materia.

Conocer los objetivos educacionales.

Conocer el grupo al cual se administrará la prueba.

Poseer gran facilidad de comunicación. No sólo deberá conocer el significado de las palabras, sino que tendrá que utilizarlas con gran precisión, para evitar falsas interpretaciones.

Conocer las técnicas apropiadas para escribir preguntas y los diferentes tipos de preguntas, características, ventajas y defectos.

Resumiendo, el proceso de escribir preguntas para una prueba no es simple. No todos los individuos poseen las habilidades para ello y, aún aquellos que las poseen, encuentran que su éxito varía con la cantidad de energía e el tiempo que pueden dedicarle a la tarea. Es necesario reconocer las dificultades que están implicadas en la construcción de una prueba para mejorar su calidad.

Durante largo tiempo se ha mantenido la polémica, tanto entre los especialistas en evaluación, como entre los profesores, acerca de los tipos de preguntas que deberán ser incluidos en sus pruebas. La evaluación educacional actual pone énfasis, no tanto en las características formales de las diversas clases de preguntas y problemas, si no en la objetividad, precisión y relevancia del instrumento de medida. De acuerdo a este criterio, la calidad de la prueba dependerá de la importancia adecuada que se de en ella a los diferentes aspectos del rendimiento y a la construcción cuidadosa de las preguntas, sin tener una influencia decisiva en este aspecto, el tipo de problema incluido. Por otra parte, cuando el puntaje de un examen revela las impresiones privadas, subjetivas, difíciles o imposibles de verificar del examinador, su significado es deficiente y no representa ninguna utilidad para el alumno ni el profesor.

Sin embargo, me parece de interés en este momento, analizar algunas de las diferencias existentes entre los dos principales tipos de preguntas utilizados actualmente en la evaluación del aprendizaje: preguntas objetivas de selección múltiple y preguntas tipo ensayo; un análisis de estas diferencias permitirá lograr una mejor comprensión de la forma en que la antigua discusión sobre el uso de estas preguntas ha sido resuelta.

En la pregunta tipo ensayo se proporciona una respuesta. El sujeto no sólo debe recordar hechos importantes para responderla, sino que debe relacionarlos y organizarlos en forma lógica, presentándolos en una composición escrita.

En la pregunta objetiva el examinando debe seleccionar la respuesta entre varias alternativas. Para efectuar este proceso deberá reconocer la respuesta apropiada y además deberá tener un amplio conocimiento de la materia, para poder descartar los distractores. Una pregunta objetiva bien construida exige al sujeto pensar en forma original antes de encontrar la solución adecuada.

En la vida diaria es necesario efectuar elecciones de todo tipo. Hacer una elección precisa es, a veces, más importante y difícil que emitir una afirmación sobre determinado problema.

Estudios de correlación efectuados en un mismo grupo de estudiantes que fueron sometidos a los dos tipos de pruebas sobre la misma materia, es decir, una en que tenían que pensar y escribir la respuesta y otra en que sólo debían elegir la mejor entre cinco opciones, demostró que ambas pruebas medían idénticos aspectos del conocimiento, tanto en relación a la materia de estudio, como a la utilización de habilidades y destrezas.

La prueba tipo ensayo tiene pocas preguntas de tipo general, que exigen una respuesta extensa. La prueba objetiva está constituida por muchas preguntas que requieren respuesta breve.

El mayor número de preguntas que puede ser incluido en una prueba objetiva representa una ventaja, al convertir la prueba en una muestra confiable de los conocimientos que se desea medir.

Mientras mayor sea el número de elementos independientes en la muestra de lo que se mide, mejor se podrá identificar el rendimiento en los diversos aspectos. Las respuestas a una prueba de ensayo compleja encierra varios elementos separados, pero, por lo general, ellos son tratados en forma más o menos integrada, tanto por el alumno como por el corrector.

La probabilidad de que un mal alumno obtenga un puntaje más alto que un buen alumno, disminuye al aumentar el número de preguntas de una prueba. En un examen de 100 preguntas hay muy poca probabilidad de obtener un buen puntaje habiendo estudiado sólo algunas materias; lo contrario puede suceder con una prueba de tres o cuatro preguntas.

En las pruebas de ensayo el alumno ocupa la mayor parte del tiempo pensando y escribiendo. En los tests objetivos los examinandos se dedican principalmente a leer y pensar.

Tanto las destrezas para leer como para escribir no son factores importantes para determinar el puntaje de los alumnos en una prueba. Luego, no podemos asegurar que ninguno de los dos tipos de pruebas comentados constituya una medida pura de los conocimientos, sin estar contaminada por destrezas de expresión o interpretación.

Sin embargo, podemos hacer una división de las rutinas de razonamiento que siguen los estudiantes al responder las diferentes formas de preguntas. Si medimos en unidades de palabras, la lectura constituye un proceso mucho más rápido que la escritura. Un estudiante típico posee una velocidad de

lectura diez veces mayor que la de escritura. Si las preguntas de una prueba objetiva contienen el mismo número de palabras que una prueba de ensayo, el alumno podría disponer de mayor tiempo para reflexionar durante el examen. Esto no nos permite afirmar que las pruebas objetivas sean mejores, pero sí, constituye un argumento contra aquellos que las atacan.

Es fácil preparar una prueba de ensayo, pero es difícil corregirla. Escribir preguntas objetivas es tedioso y difícil, pero no constituye ninguna dificultad corregirlas. Este hecho es importante y decisivo para determinar la situación en que debe emplearse cada tipo de pregunta. Si un examen va a administrarse a muchos sujetos y se tiene tiempo suficiente para escribir las preguntas, es preferible el examen con preguntas objetivas. Pero si la prueba tiene un uso limitado, con pocos estudiantes y el profesor no está obligado a entregar los resultados en un plazo breve, puede elegir preguntas tipo ensayo.

La prueba tipo ensayo da mayor libertad al alumno para expresar su respuesta y mayor libertad al corrector para corregir guiándose por su opinión personal.

Las pruebas objetivas dan mayor libertad al constructor; el comportamiento del examinado y del corrector está controlado.

Considerado desde este punto de vista, el test de ensayo podría caracterizarse como una prueba proyectiva en la cual intervienen con mayor fuerza los intereses y valores de los alumnos que los intereses del constructor para determinar la respuesta.

La palabra libertad y su concepto es un valor específico, no universal; en algunas situaciones es muy valiosa, pero en otras resulta perjudicial. En medición educacional, la libertad es enemiga de la precisión.

Mientras más preciso sea el resultado de nuestra medición, más detallada será la descripción y mayor la standarización del proceso de medición; todas las variables que afectan el resultado deben ser controladas. Desde este punto de vista, la prueba objetiva representa un avance técnico en la medición educacional.

La prueba de ensayo, al permitir al alumno elegir su respuesta, permite a los diferentes estudiantes competir en dimensiones diversas, aumenta la dificultad de comparación entre los resultados; la libertad del corrector le permite atender a lo que ya sabe del alumno; una respuesta correcta de un mal alumno tiende a ser subestimada, mientras que una respuesta mediocre de un buen alumno es evaluada más alto de lo que merece.

Se ha dicho que los tests objetivos son simplistas, que la vida no presenta problemas tan fáciles como "escoger" una respuesta entre cinco posibilidades y que los alumnos adquieren la noción errada de que existe una sola respuesta correcta a cada problema. Consideremos que si bien es cierto que una prueba tipo ensayo, que permite diversas respuestas correctas, valorizando las razones de ellas es valiosa, un estudiante que lucha para elegir la mejor entre varias soluciones plausibles y que después descubre que su elección no fue la mejor también aprenderá, debido a la complejidad del problema presentado.

En las pruebas objetivas la tarea del alumno y las bases sobre las cuales se juzgarán los resultados están establecidas en forma más clara que en las pruebas de ensayo.

No debemos olvidar, sin embargo, que las preguntas de ensayo pueden hacerse mucho más precisas acompañándolas de explicaciones detalladas e instrucciones explícitas. La construcción de estas preguntas también puede ser guiada por instrucciones apropiadas, dirigiendo al alumno hacia una situación específica y permitiéndole luego, reaccionar dentro de ciertos límites.

Tanto las pruebas de ensayo como las objetivas pueden ser usadas para medir todo tipo de rendimiento educacional que pueda ser medido mediante pruebas escritas.

Ambos tipos de pruebas pueden utilizarse para motivar a los alumnos a estudiar, teniendo como objetivo principal la comprensión de principios, la organización o integración de ideas y la aplicación de conocimientos a la solución de problemas.

Resumiendo, las pruebas tipo ensayo se utilizan cuando:

El grupo al que se administra es pequeño y la prueba no se repetirá.

El profesor desea motivar y premiar el desarrollo de la habilidad de expresión escrita de los alumnos.

El profesor está más interesado en explorar las actitudes del estudiante que en medir conocimientos.

El profesor tiene más confianza en su habilidad para corregir preguntas de respuesta abierta o ensayo que para escribir preguntas objetivas.

El tiempo disponible para preparar la prueba es más corto que el tiempo para corregirla.

Se utilizan preguntas objetivas cuando:

El grupo de examinados es numeroso.

Es necesario obtener puntajes tan confiables como sea posible.

Es esencial la imparcialidad en la evaluación.

El profesor tiene más confianza en su habilidad para construir pruebas objetivas que para juzgar respuestas de ensayo.

El plazo para entregar los resultados es más breve que el necesario para construir rápidamente la prueba.

Ambas clases de preguntas aporta importantes contribuciones a la medición educacional. Aunque difieren en su forma, son intercambiables. Más importantes que decidir cuál forma se utiliza es utilizarlas bien.

Análisis de las preguntas

¿Cómo se logra precisar la calidad de una prueba? El estudio de los factores que deben ser considerados al juzgar la calidad de una prueba constituye una etapa fundamental en la construcción de los instrumentos de medida y puede efectuarse, en parte mediante un cuidadoso análisis lógico y de contenido de la prueba misma y de su correspondiente tabla de especificaciones; otros aspectos, sin embargo, necesitan un análisis estadístico de los puntajes obtenidos y de cada pregunta en particular.

Una vez que se ha construido un número apropiado de preguntas, que éstas han sido criticadas por especialistas en la materia y en construcción de pruebas, deben ser administradas experimentalmente a una muestra de examinados.

Esta muestra debe ser tan representativa del grupo al que se aplicará la forma definitiva de la prueba, como sea posible. Esta aplicación experimental es forzosa en toda prueba a nivel nacional o de cuyos resultados deban derivarse decisiones de fundamental importancia para los examinados. En aquellas pruebas de rendimiento, elaboradas por el profesor de un curso en colaboración con sus ayudantes, el estudio de las preguntas puede hacerse inmediatamente después de su administración y corrección, omitiéndose el "pre-test" o aplicación experimental.

Existen diferentes procedimientos estadísticos para analizar, tanto las preguntas de un test, como la calidad de la prueba en su totalidad. Estas técnicas pueden ser utilizadas, en forma simplificada, por el personal docente de la universidad; nuestra experiencia en Chile, nos ha demostrado, sin embargo, que pueden colaborar con el personal académico, especialistas en medición educacional, los que efectúan especialmente el estudio de los exámenes de admisión a la universidad o aquellos exámenes de los cuales dependen decisiones de fundamental importancia para los examinados.

El análisis estadístico se aplica más fácilmente a las pruebas con preguntas objetivas; puede también ser aplicado, al menos en parte, a los tests tipo ensayo. Toda la información proporcionada por el análisis estadístico de las preguntas es una información detallada, objetiva y cuantitativa.

Los propósitos del análisis estadístico de las preguntas podrían resumirse así:

- 1) Identifica las preguntas defectuosas o débiles para mejorarlas, si es posible (preguntas ambiguas, distractores que no funcionaron, etc.).
- 2) Determina la dificultad de cada pregunta, para seleccionar aquellas que muestren un nivel apropiado.
- 3) Determina el nivel discriminativo de cada pregunta, para elegir aquellas que contribuyan a lograr un instrumento de medición eficiente.
- 4) Ayuda a mejorar la confiabilidad de la prueba, con la selección de preguntas apropiadas.

Sin embargo, se considera que el dato cuantitativo no basta como sustituto del estudio analítico de las preguntas basado en la experiencia y conocimientos del especialista y del profesor de determinada materia. Ambos estudios se complementan y deben ser interpretados uno en base al otro.

Validez y confiabilidad

La evaluación educacional basada en la medición exige de una prueba, además de una elaboración científica de preguntas, determinado nivel de validez y confiabilidad.

La característica que más afecta el valor de una prueba es su validez, la que ha sido definida por Lindquist como "la exactitud con que el instrumento mide lo que se desea medir". Esta definición implica que para determinar la validez de una prueba, debemos "comparar la realidad de lo que mide con el concepto ideal de lo que debiera medir". Una definición más operacional, aunque más estrecha de la misma idea es la siguiente: "La validez de una prueba es una estimación de la correlación entre los puntajes obtenidos en la prueba y los puntajes reales de un criterio".

La validez es alta si el test mide lo que se trata de evaluar, es decir, da la información que se desea. No importa la perfección que la prueba revele en otros aspectos; una prueba que mide un objetivo equivocado, no tiene valor.

La validez es materia de grado. No representa una característica absoluta de existencia o no existencia. No hablaremos de pruebas válidas o no válidas, si no que de pruebas más o menos válidas.

Esta característica no es intrínseca a la prueba misma. Depende del propósito para el cual se utiliza, del grupo y forma en que se administra y de las técnicas con que se corrige. En lugar de preguntar: ¿Cuál es la validez de esta prueba?, debiéramos preguntar: ¿Cuál es la validez de los puntajes de esta prueba cuando se usa de modo específico, para un objetivo específico y en una población de sujetos específica? La mayor parte de los tests construidos por el profesor universitario se construye teniendo presente un grupo y un propósito determinado y sólo serán útiles para grupos semejantes y propósitos idénticos.

Según el propósito con que se utilizará la prueba, su validez se obtendrá empíricamente, mediante técnicas estadísticas apropiadas, o a través de un proceso de análisis de su contenido o de los conceptos teóricos implicados en su elaboración.

Lo más difícil, en los estudios de validación empírica, es la obtención de un criterio que pueda ser sometido a medición objetiva. El criterio constituye un registro o conjunto de datos del comportamiento de los sujetos o de los resultados obtenidos por ellos. Ejemplos de estos criterios de validación podrían ser: las opiniones de profesores, la cantidad de materia aprendida, las notas obtenidas, etc.. Los criterios pueden ser más de uno y deben abarcar todos los aspectos del estudio.

Todos los tipos de validez requieren definición, explícita o por implicación, de los rasgos que la prueba en estudio trata de medir. Ya ha anotado Cook que una característica de toda propiedad susceptible de ser medida, es que ésta pueda ser definida claramente y sin ambigüedad. Un rasgo, para poder ser definido operacionalmente, debe serlo en términos de la tarea que diferenciará a los que lo poseen en alto grado, de los que lo tengan en grado menor. Por ejemplo: no existe mejor medio de aclarar lo que tratamos de decir con la expresión "rendimiento en álgebra o química", que la descripción de cómo mediríamos el grado en que una persona posee este rendimiento.

Por último, un importante componente de la validez es la relevancia. La relevancia no puede medirse en forma estadística, basándose en datos experimentales; debe ser construida con la prueba. Para la mayoría de las pruebas construidas por el profesor, es materia de análisis lógico y de razonamiento. Las decisiones que tome en cada paso del proceso, determinan la relevancia del examen. Si las preguntas del test son todas relevantes, si todas requieren demostración del dominio de algún aspecto esencial del curso y si

representan una muestra proporcional de aquellos aspectos esenciales, podemos considerar que la prueba es válida.

El análisis científico de un instrumento de medición educacional implica, además, como hemos dicho, el cálculo de su confiabilidad. Esta constituye, para la mayor parte de las pruebas de conocimiento, el índice estadístico de mayor valor. Una manera de aclarar este concepto, es compararlo con el de validez:

Se entiende por confiabilidad de un examen, la consistencia con que un conjunto de puntajes mide; validez, por otra parte, es la exactitud con que un conjunto de puntajes mide lo que debe medir.

Si todos los puntajes obtenidos fueran exactos, sin estar viciados por errores de muestreo de las preguntas, ansiedad o fatiga de los examinandos, etc., el coeficiente de confiabilidad sería 1,00. Nunca se logra este grado de perfección: el error es imposible de evitar en las mediciones, pero el objetivo de los especialistas es reducir estos errores a un mínimo.

Una prueba confiable, es decir que mide en forma precisa, no es necesariamente válida. La confiabilidad es una condición necesaria para la validez, pero no es una condición suficiente. El autor de una prueba de confiabilidad alta puede haber obtenido éxito en medir con gran precisión, algo no importante.

La confiabilidad se traduce en un "coeficiente de confiabilidad" que es una estimación del coeficiente de correlación entre un conjunto de puntajes de un test administrado a un grupo de sujetos y los puntajes de otro test equivalente, administrado a los mismos examinandos. Este coeficiente nos da a conocer la confianza que podemos tener en la medida obtenida.

Cómo puede el profesor aumentar la confiabilidad de sus pruebas? Tomando en cuenta los factores que la afectan y que él puede controlar. Esto implica escribir, revisar y seleccionar las preguntas de modo que discriminen entre los alumnos de alto y bajo rendimiento en la habilidad medida; exige, también, elegir preguntas de dificultad apropiada e incluir tantos ítem, como sea posible responder en el tiempo disponible. Por último, implica una administración en condiciones adecuadas.

Otros medios utilizados en la evaluación del aprendizaje -

La evaluación de los alumnos universitarios, sin embargo, a pesar de fundamentarse básicamente en los instrumentos de medición elaborados en forma científica, incluye también otros medios de aportar información acerca de

los cambios producidos por la enseñanza. En efecto, es necesario en algunos cursos evaluar actitudes, aptitudes, intereses o hábitos. Hasta el momento, a pesar de existir algunas escalas o instrumentos contruidos para este efecto, éstos no están al alcance de todos los profesores y, aunque lo estuvieran, por lo general no se encuentran standarizados o adaptados de acuerdo a las técnicas exigidas. La observación directa constituirá, en estos cursos, el medio más apropiado para conocer el desarrollo de las diferentes características de la personalidad.

La observación ha sido llamada el "instrumento de evaluación psicológica más antiguo": en efecto, es un instrumento como cualquier prueba y, en consecuencia, debe ser utilizado en condiciones que permitan un máximo control de la situación. Esta, a su vez, debe ser standarizada, debe constituir el mismo estímulo para todos los alumnos y permitir estudiar lo mismo a todos los observadores. La observación, así entendida, provee una muestra real del comportamiento del individuo y permite, por lo tanto, evaluar la conducta espontánea del alumno; además, en la misma forma en que la prueba motiva a estudiar, la observación motiva a cambiar de conducta, especialmente si se realiza de modo continuo.

Para obtener el máximo de objetividad, la observación debe ser estructurada, es decir debe seguir un plan determinado y tener un propósito específico: debemos, también, exigirle la mayor precisión posible. Existen técnicas que permiten aumentar la confiabilidad de este método de evaluación y que deben ser conocidas por todos los profesores que lo utilicen.

Según el propósito de la observación, ésta se resume, se traduce, se anota y se comunica en distintos tipos de escalas; no es correcto considerar estas escalas como instrumentos de evaluación, ya que su objetivo es solamente anotar lo observado y restringir las observaciones a determinados aspectos del comportamiento. No son instrumentos que se administren al individuo, ya que no constituyen un estímulo para él.

Sistemas de notas -

Los resultados de la evaluación educacional se traducen, generalmente, en puntajes o notas. Por lo general, estos puntajes se utilizan sin comprensión de sus limitaciones. Un puntaje por sí mismo, no tiene significado; debe ser interpretado en relación a algún grupo de referencia o a ciertos standards pre-establecidos.

Calificar una prueba constituye un problema complejo y difícil. Desde algunos aspectos es aún más difícil que construir un buen test; por esto, es necesario estudiar cuidadosamente la forma apropiada de otorgar los puntajes interpretarlos y utilizarlos.

Las notas se han convertido en materia de controversia. Los conceptos predominantes en el campo educacional, en épocas diferentes de su evolución han influido en el valor que se les ha conferido. Mientras en la tercera y cuarta década de este siglo, poniendo énfasis en el desarrollo de la personalidad y en la democracia en el campo educacional, se les criticó por constituir una presión competitiva que medía con un standard común a todos los alumnos, en la década actual se pide una evaluación más rigurosa de los conocimientos, de acuerdo al actual sistema de valores imperante.

Es necesario que toda facultad posea o desarrolle un sistema de notas. Se ha dicho que los sistemas de notas son especialmente materia de definición; esta definición deberá ser conocida tanto por los profesores como por los alumnos.

Si no se efectúan estas definiciones en forma apropiada y si no se otorgan las calificaciones de acuerdo a ellas, dejarán de tener valor como tales. Un sistema de notas es básicamente un sistema de comunicación. Implica la utilización de símbolos especializados cuyo significado debiera ser preciso y uniformemente comprendido. Sólo cuando las notas poseen el mismo significado para todos los que las utilizan, es posible que sirvan a los propósitos de la evaluación.

La forma en que un profesor, un departamento o toda una facultad califica a los estudiantes, tiene importancia para otros profesores, otros departamentos y otras facultades. Antes de colocar sus notas, el profesor deberá hacerse las siguientes preguntas:

¿Deberá poner su nota de acuerdo al programa, a la cantidad de materia, o en relación al rendimiento total del grupo?

Estas notas, ¿se otorgarán por rendimiento en el aprendizaje o por las actitudes, el esfuerzo, el carácter, etc.?

¿Se otorgará una sola nota o varias, por diversos aspectos del rendimiento?

¿Se otorgarán notas por la cantidad de conocimientos poseídos o por el aumento de ellos en determinado período de tiempo?

Todas y cada una de las respuestas a estas preguntas influirán en la calidad de las unidades de medida, en su confiabilidad y su validez.

En resumen, la enseñanza y la evaluación no deben ni pueden ser consideradas como alternativas independientes una de otra o mutuamente excluyentes; por el contrario, constituyen aspectos íntimamente relacionados del proceso educacional total. El valor de una prueba depende de su calidad y de la forma en que se la utiliza; exámenes contruidos de acuerdo a las técnicas apropiadas, aportan contribuciones valiosas al aprendizaje; pruebas deficientes o administradas en forma defectuosa no influyen en este proceso o, aún más, pueden influir negativamente.

Muchos excelentes profesores parecen demostrar indiferencia o desprecio por construir exámenes que permitan una adecuada evaluación de los conocimientos de sus alumnos. Considero que esta actitud y el vacío que generalmente existe entre nuestras aulas y los especialistas en construcción de pruebas, es causado por un problema de comunicación: los especialistas utilizan un idioma propio, que parece ubicar sus conocimientos más allá del alcance del profesor. Los catedráticos debieran confiar más en su experiencia y utilizar las técnicas relativamente simples del análisis estadístico, lo que, unido al esfuerzo que representa la construcción de preguntas apropiadas, les permitirá construir instrumentos de medición adecuados a la evaluación de sus cursos.

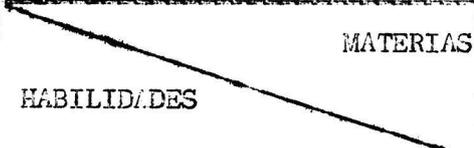
ESQUEMA PARA LA CONSTRUCCION DE UN EXAMEN DE BIOLOGIA
(Traducido de "Test construction" de Dorothy Adkins Wood)

OBJETIVOS	I - TENER CONOCIMIENTO Y COMPRENSION DE:			II - RAZONAMIENTO CIENTIFICO EVIDENCIADO EN LA HABILIDAD PARA:			
	A. VOCABULARIO SOBRE EL TEMA	B. CONCEPTOS Y PRINCIPIOS DE LA BIOLOGIA	C. LECTURAS EN RELACION AL TEMA	A. RECONOCER Y RESOLVER PROBLEMAS	B. RECONOCER HIPOTESIS Y SELECCIONAR METODOS PARA COMPROBARLAS	C. EVALUAR EN FORMA CRITICA PROCEDIMIENTOS, DATOS, CONCLUSIONES EXPERIMENTALES	D. JUZGAR SITUACIONES REALES
MATERIAS	7% de los items	35% de los items	8% de los it.	15% de los it.	5% de los items	15% de los items	15% de los it.
Características comunes a todos los seres vivos	----- 4% DEL TOTAL DE ITEMS -----			----- 3% DEL TOTAL DE ITEMS -----			
Conservación del individuo							
1. Producción de alimentos y su utilización							
2. Transporte y excreción	----- 17% DEL TOTAL DE ITEMS -----			----- 18% DEL TOTAL DE ITEMS -----			
3. Coordinación y adaptación							
Conservación de las especies							
1. Reproducción							
2. Herencia	----- 12% DEL TOTAL DE ITEMS -----			----- 12% DEL TOTAL DE ITEMS -----			
Historia de la vida sobre la tierra	----- 5% DEL TOTAL DE ITEMS -----			----- 4% DEL TOTAL DE ITEMS -----			
Interrelaciones							
1. Relaciones ecológicas							
2. Parasitismo y Enfermedad	----- 12% DEL TOTAL DE ITEMS -----			----- 13% DEL TOTAL DE ITEMS -----			
3. Clasificación							
	----- 50% DEL EXAMEN -----			----- 50% DEL EXAMEN -----			

ANALISIS DE PREGUNTAS

1. Arreglar las pruebas corregidas o las "hojas respuestas" en orden de puntajes, de las más altas a las más bajas.
2. Separar 2 sub-grupos de pruebas, el grupo superior (27% del grupo total) y el grupo inferior (27%).
3. Contar las veces que cada alternativa de cada pregunta, fue elegida en las pruebas del grupo superior. Hacer lo mismo, en forma separada, para las pruebas del grupo inferior.
4. Anotar estos dos resultados junto a las respuestas pertinentes, en una copia de la prueba.
5. Sumar las dos cantidades obtenidas para la clave. Dividir el resultado de esta suma por la suma máxima posible (la suma del número de pruebas en ambos grupos). Expresar este cociente como un porcentaje, es decir, multiplicar la fracción decimal por 100.
El resultado es el índice de dificultad de la pregunta.
6. Restar el número de respuestas correctas del grupo inferior de las del grupo superior. Dividir esta diferencia por el número de pruebas contenido en el grupo superior o inferior (27%). Este cociente, expresado como fracción decimal, es el índice de discriminación.

ESQUEMA DETALLADO DE MATERIAS Y HABILIDADES A MEDIRSE EN UN TEST DE BIOLOGIA
 Y CIENCIAS

DISTRIBUCION DE ITEMS 	QUIMICA	FISICA	ASTRONOMIA	GEOLOGIA	METEOROLOGIA	HISTOLOGIA	BOTANICA	ZOOLOGIA ANATOMIA Y FI- SIOLOGIA HUMANA	BIOLOGIA E HI- GIENE HUMANA	ECOLOGIA	HERENCIA Y EVOLUCION	TOTAL
	Comprensión de principios y conceptos científicos básicos											
Habilidad para distinguir conceptos básicos de aquellos que están fuera del tema o son inapropiados.												
Habilidad para anticipar y diagnosticar conceptos que serán fáciles o difíciles para los alumnos.												
Habilidad para seleccionar y construir demostraciones que contribuyan a una enseñanza efectiva.												
Habilidad para reconocer y utilizar fuentes de información apropiadas.												
Habilidad para dar explicaciones claras de principios y conceptos científicos.												
Habilidad para aplicar principios y conceptos científicos a las experiencias de la vida diaria.												
Habilidad para evaluar los conocimientos de los alumnos.												
Habilidad para estimular y guiar a los alumnos.												
T O T A L												

- Traducido del Folleto "ETS Builds a Test" del Educational Testing Service p.5
- Esta tabla de doble entrada es una reproducción de aquella utilizada para "Biology and General Science Test in the National Teacher Examinations".
- Este esquema previo a la construcción del test, muestra las habilidades y materias que deben incluirse en el test, en cada cuadro se inscribe además, el número de preguntas de cada tipo que se construirán.

Material seleccionado y traducido por el Instituto de Investigaciones Estadísticas de la Universidad de Chile para sus Seminarios de Construcción de Pruebas.

PRUEBA

Probado en:	Utilizado en:									
Forma										
Número										
Números absolutos	N	A	B	C	D	E	N/A	O	p	clave r _b
Porcentajes										
Promedio de puntajes en cada grupo										
Transformación lineal										

040/69/PR