

Unidad Académica de Educación Química



Diseño de Pruebas Objetivas

Emy Soubirón y Soledad Camarano

Junio, 2006

INDICE

CONTENIDO	PAGINA
1.- Selección de contenidos de la prueba.....	3
2.- Redacción de la prueba	5
3.- Corrección y puntuación de las pruebas	16
4.- Presentación de la prueba	17
5.- Críticas a las pruebas objetivas	18
6.- Ventajas de las pruebas objetivas	18
7.- A modo de resumen: ¿Qué pasos implica el diseño de una prueba objetiva?	19
8.- Bibliografía.....	21
9.- Anexo: Redacción de ítems de opción múltiple.....	22

PRUEBAS DE EVALUACIÓN OBJETIVAS

Las pruebas de múltiple opción

Las pruebas de múltiple opción u objetivas se componen de un conjunto de preguntas claras y precisas que requieren por parte del alumno, una respuesta breve, en general limitadas a la elección de una opción ya proporcionada. El término *objetivas* hace referencia a las condiciones de aplicación de la prueba así como al tratamiento y posterior análisis de los resultados pero ello no implica una mayor objetividad en la evaluación del rendimiento del estudiante.

Algunos aspectos a considerar en la elaboración de una prueba objetiva son:

- La selección de los contenidos
- La redacción de las preguntas o ítems
- La corrección y puntuación
- Presentación de la misma.

1.- Selección de contenidos de la prueba

En la selección de los contenidos, todos los que resulten relevantes en relación al dominio a medir deben de estar especificados de forma operativa y referidos a los objetivos buscados así como acorde a las capacidades de los estudiantes.

Se debe de establecer el nivel en que serán medidos esos contenidos así como el peso que tendrán en relación al conjunto de la prueba. Para ello se elabora una ***tabla de especificaciones*** que consiste en una tabla de doble entrada que indica los contenidos y competencias (destrezas que se ponen en juego ante el planteo de situaciones nuevas) en las columnas y filas correspondientes. En cada intersección se presenta un *objetivo evaluable* representado por los ítems propuestos en cada caso, cuyo número se

especifica en la tabla. Ésta debe ser confrontada con la tabla de resultados de la prueba una vez que la misma ha sido corregida

Un ejemplo de tabla de especificaciones, extraída de de Fleitas y otros (2002), es la que se muestra en el presente trabajo.

Se trata de una prueba de 30 ítems en total.

Los **contenidos** a evaluar (columnas de la tabla) son 4: sistemas materiales, los seres vivos y su ambiente, el ser humano y la salud, el procedimiento científico. Se observa que a los temas "sistemas materiales" y "procedimiento científico" les corresponden 11 y 8 ítems respectivamente, en tanto que a los restantes les corresponden 5 y 6.

En cuanto a las **competencias** a evaluar en los estudiantes (filas), la prueba está dirigida a evaluar 3 tipos de destrezas: nivel básico (identificación, reconocimiento y reproducción de información), nivel intermedio (interpretación de información) y nivel superior (producción de información). Se dedica aproximadamente un tercio de la prueba a cada tipo de destreza.

Evaluación Censal de Aprendizajes en Ciencias de la Naturaleza en Noveno Grado de Escuelas Rurales, año 2002.

Contenidos Competencias	Sistemas materiales	Los seres vivos y su ambiente	El ser humano y la salud	Procedimiento científico	Nº Total de ítems
Identificación, reconocimiento y reproducción de la información	2	1	5	2	10
Interpretación de la información	5	2	1	3	11
Producción convergente de la información	4	2	0	3	9
Nº total de ítems	11	5	6	8	30

Fuente: Programa MEMFOD

Las normas más comunes aplicadas a la selección de contenidos para la elaboración de una prueba podrían sintetizarse en las siguientes:

- En la prueba deben estar considerados los contenidos que se quieren medir los cuales deben ser explícitos y conocidos a priori por los estudiantes.
- Las preguntas deben de referirse a los aspectos más importantes del dominio, conteniéndolos.
- El número de preguntas o ítems para cada aspecto del dominio deberá ser proporcional a la importancia de los contenidos a ser evaluados.
- Se deben conocer, previo a la elaboración de la prueba, cuales son los objetivos alcanzados así como las capacidades o competencias desarrolladas que han de ser evaluadas.
- Las preguntas deben ser presentadas de modo tal que impliquen una dificultad creciente. Se suelen colocar al inicio las más sencillas, luego un incremento de dificultad hasta el máximo y finalmente algunas cuestiones de menor dificultad en poca cantidad.
- Conviene que la prueba sea sometida a la opinión de expertos (otros colegas de la asignatura, especialistas en la temática, pedagogos, etc.)

2.- Redacción de la prueba

Antes de considerar la forma de redacción propiamente dicha, conviene aclarar que se entiende por *pregunta*. Mateo (2000:74) establece "*suele entenderse por pregunta una declaración, asunto o tema sobre el que se va a trabajar.*" También puede utilizarse el término *ítem* que puede expresarse no solo en forma interrogativa, como la pregunta, sino que puede ser una afirmación, una instrucción que se le da al estudiante, una orden para realizar una tarea, etc.

Algunas características de los ítems:

- ✓ Deben ceñirse a las competencias y contenidos preestablecidos en la tabla de especificaciones.
- ✓ Deben ser independientes entre sí y exhaustivos en cuanto a la información necesaria para su resolución.
- ✓ La respuesta de un ítem no puede ni debe ser condición para la resolución de ninguno de los siguientes ítems.
- ✓ No deben de incluir nombres de ficción o marcas de productos usados en los medios masivos de comunicación.
- ✓ En una prueba deben de plantearse ítems de diversos grados de dificultad.

El ítem consta de una **base o raíz** expresada en forma aseverativa o interrogativa, que presenta la situación problema y una serie de alternativas de respuesta. Una o varias de estas alternativas, van a ser las correctas (**clave/s**).

Con relación a la base o raíz la misma debe cumplir con ciertas características tales como:

- ✓ Claridad, brevedad e inteligibilidad.
- ✓ Ser escrita en lenguaje directo y comprensible para el respondente.
- ✓ Referirse a un solo contenido.
- ✓ Ser expresada, de preferencia, en modo positivo.
- ✓ Cuando impliquen la inclusión de información textual, la misma se debe de encomillar y acompañar de la referencia a la fuente para verificar su fidelidad.

Las respuestas que son plausibles pero no correctas, se denominan **distractores**. Los distractores surgen de los errores más comunes que presentan los estudiantes en sus respuestas, asociados a la cuestión planteada y no son afirmaciones "solo de relleno", poco creíbles. Deben de tener todos aproximadamente la misma extensión. La respuesta correcta o

clave debe presentarse aleatoriamente entre las alternativas cuyo número ideal es de cuatro a cinco, nunca menos de tres, así se pueden controlar aceptablemente los errores del azar. No debe de destacarse de los distractores por tener una redacción más amplia o enfatizada. Pueden usarse distractores del tipo “ninguna de las anteriores” o “no hay datos suficientes para contestar”, sin abusar de su uso.

Cada ítem debe analizarse particularmente a los efectos de determinar sus características y facilitar su ubicación en una base de datos. Una posibilidad es realizar el siguiente cuadro **(perfil del ítem)** adaptado del Programa de Evaluación de Aprendizajes de ANEP, año 2002.

Título del ítem:		
Competencia a evaluar:		
Contenido a evaluar:		
Base o raíz del ítem:		Proceso a desarrollar para lograr la respuesta:
		Justificación de distractores:
Opciones de respuestas:	A	
	B	
	C	
	D	
	E	
Grado de dificultad a juicio del experto:		Observaciones:
Nombre del redactor:		Firma:
Fecha de redacción:		Nivel de aplicación:

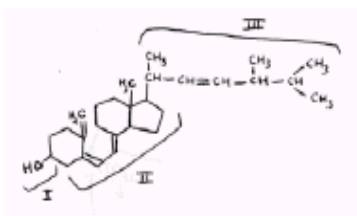
Según Mateo (2000) existen gran cantidad de posibles ítems, algunos de los cuales, se ilustran a continuación. *Los ejemplos empleados corresponden a diversas áreas del conocimiento así como a diferentes niveles educativos.*

- a. **De selección simple:** se caracteriza por presentar una sola respuesta correcta y una serie de distractores.

Ejemplo 1

Se muestra la estructura del calciferol, o vitamina D₂, con tres regiones de la molécula indicadas como I, II y III.

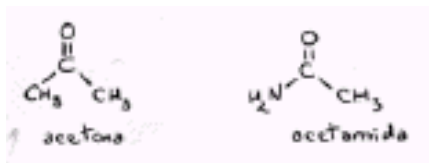
Ud. razona que:



- a. las regiones I, II y III son todas hidrofílicas
b. las regiones I y II son hidrofílicas; la región III es hidrofóbica
c. la región I es hidrofílica; las regiones II y III son hidrofóbicas
d. las regiones I y III son hidrofílicas; la región II es hidrofóbica
e. las regiones I, II y III son todas hidrofóbicas

Ejemplo 2

Considere las moléculas de acetona y acetamida.



Ud. razona que una molécula de acetona puede formar enlaces de hidrógeno con:

- i. moléculas de agua
ii. moléculas de acetamida
iii. otras moléculas de acetona

- a. i solamente
- b. ii solamente
- c. iii solamente
- d. i y ii solamente
- e. i, ii y iii

Fuente: <http://mail.fq.edu.uy/~biol/> (Introducción a las Ciencias Biológicas, Facultad de Química, Udelar)

- b. **De selección múltiple:** hay más de una respuesta verdadera entre diferentes alternativas. *Debe aclararse en la proposición esta situación.*

Ejemplo

¿A qué nivel (es) de estructura de proteínas contribuyen los enlaces de hidrógeno?

- a. primario
- b. secundario
- c. terciario
- d. cuaternario

-
- c. **De selección de la mejor respuesta:** todas las proposiciones son verdaderas y el estudiante debe determinar cual es la mejor.

Ejemplo

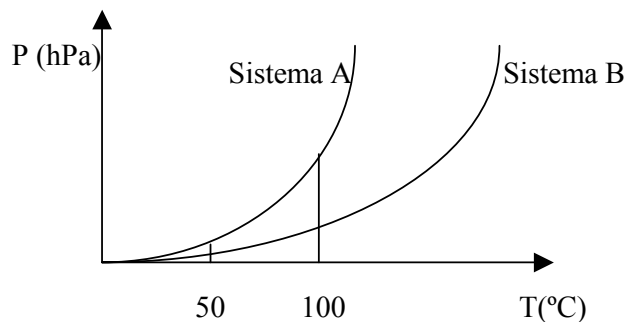
Cuál de las siguientes expresiones ilustra mejor la relación entre presión y volumen en un gas ideal:

- a.- La presión y el volumen varían en forma inversamente proporcional en determinadas condiciones.
 - b.- La presión y el volumen varían en forma inversamente proporcional si la cantidad de sustancia no cambia.
 - c.- La presión y el volumen varían en forma inversamente proporcional si la temperatura y la cantidad de sustancia son constantes.
 - d.- La presión y el volumen varían en forma inversamente proporcional si la temperatura es constante.
-

- d. **De selección de la respuesta incorrecta:** en este caso el estudiante debe seleccionar la respuesta que no es correcta y los distractores son respuestas correctas. *Se debe aclarar muy bien en las instrucciones que se ha de elegir la incorrecta.*
-

Ejemplo:

Considerando la gráfica de la figura de presión de vapor en función de la temperatura, para los sistemas A y B, ¿cuál de las siguientes afirmaciones es incorrecta ?



1. El sistema B tiene menor presión de vapor que el sistema A
 2. El sistema A tiene igual punto de ebullición normal que el sistema B
 3. El sistema B tiene mayor punto de ebullición normal que el sistema A
 4. El sistema A tiene mayor presión de vapor que el sistema B
-

- e. **De base común:** se aprovecha la base para generar en torno a ella, diversos conjuntos de respuestas.
-

Ejemplo

Para las preguntas 1 y 2 considere las siguientes condiciones, asociadas a problemas de la tiroides o de la transducción de las señales portadas por las hormonas tiroideas.

- i. deficiencia de yodo en la dieta (bocio por deficiencia de yodo)
- ii. estimulación excesiva de la tiroides debida a una reacción inmunológica inapropiada (enfermedad de Graves)
- iii. receptor para T3/T4 mutado, no funcional (resistencia a las hormonas tiroideas)

1 ¿En cuál(es) de las condiciones los individuos poseerán un metabolismo energético más activo que lo normal?

- a. i solamente
- b. ii solamente
- c. iii solamente
- d. i y ii solamente
- e. ii y iii solamente

2. En Uruguay, toda la sal de mesa comercial está suplementada con yoduro. ¿Qué condición(es) previene(n) esta medida?

- a. ninguna de las listadas
- b. i solamente
- c. i y ii solamente
- d. i y iii solamente
- e. i, ii y iii

Fuente: <http://mail.fq.edu.uy/~biol/> (Introducción a las Ciencias Biológicas, Facultad de Química, UdelaR)

-
- f. **De ordenamiento:** en este tipo de ítems, el estudiante se enfrenta a una serie de hechos o sucesos desordenados y los debe ordenar según un criterio previamente establecido. Algunos de los criterios más comúnmente usados son: histórico, espacial, causal, funcional, lógico, etc.

Ejemplo

En la generación de la respuesta de una célula postsináptica a un estímulo presináptico ocurren los siguientes procesos:

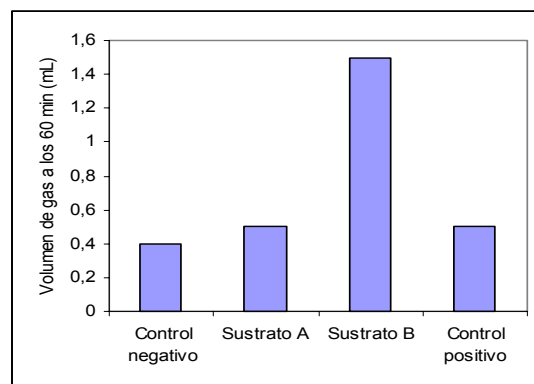
- 1.- Liberación del neurotransmisor.
- 2.- Fusión de las vesículas presinápticas con la membrana.
- 3.- Entrada de calcio a través de canales regulados por voltaje.

Ordene los eventos anteriores en la secuencia en que ocurren en el organismo.

- g. **De identificación de gráficos y localización en mapas:** se pide al estudiante que sobre un mapa o gráfico realice alguna identificación o localización.
-

Ejemplo

Se realizan una serie de experimentos para determinar la capacidad de la levadura del pan de utilizar dos sustratos (A y B). Se mide el volumen de gas desprendido y luego de 60 minutos se obtienen los siguientes datos:



¿Cuál de estas afirmaciones puede desprenderse de los datos obtenidos?

- a. El sustrato A no es fermentado por la levadura en estas condiciones
 - b. El sustrato A es fermentado a menor velocidad que el sustrato B
 - c. Los sustratos A y B no son fermentados por la levadura en estas condiciones.
 - d. No es posible concluir acerca de la capacidad de fermentación de la levadura respecto a estos sustratos.
-

- h. **De verdadero o falso:** se debe determinar la corrección o incorrección de un grupo de respuestas.

Ejemplo:

	V	F
Un electrófilo es una especie ávida de cargas positivas		
En un carbono secundario, una sustitución nucleofílica se orientará vía SN1 mediante un solvente polar		
En los alenos todos los carbonos son sp^2		
El 1-nitronaftaleno se nitrará mas rápidamente que el nitrobenceno		
El 1-nitronaftaleno se nitrará mas rápidamente que el 1-naftol		

Fuente: <http://gea.ciens.ucv.ve/~manmarti/orga/p4resuelto.pdf> (Universidad Central de Venezuela, Química Orgánica)

- i. **De emparejamiento:** se componen de dos listas (premisas y respuestas) con datos a emparejar. Pueden usarse combinaciones del tipo: fechas y sucesos o conceptos y definiciones o magnitudes y unidades, etc.

Ejemplo

Une mediante flechas los siguientes compuestos químicos y su fórmula:

- | | |
|---------------------|----------|
| 1. Sulfuro sódico | $HClO_3$ |
| 2. Oxido de cinc | Na_2S |
| 3. Ácido clórico | $AgNO_3$ |
| 4. Nitrato de plata | ZnO |
-

- j. **De problemas científico-matemáticos:** se presenta una premisa a resolver y posibles respuestas numéricas que den cuenta de esa solución. Deben ser muy bien ajustadas a los objetivos y los distractores

deben poder discriminar entre errores frecuentes asociados al proceso de resolución. Permiten comprobar la comprensión y aplicación de un conocimiento en contraste con la memorización del mismo.

Ejemplo

Un astronauta se encuentra en el espacio a 20,0 metros de su nave. La nave tiene los motores apagados y el astronauta está inicialmente en reposo respecto a ella. El astronauta tiene tres herramientas, y para acercarse a la nave va tirando herramientas en sentido contrario a la dirección en que se encuentra ésta, a razón de una herramienta cada 10,0 segundos. La velocidad de las herramientas respecto del astronauta es de 10,0 m/s. Inicialmente el astronauta y sus herramientas tienen una masa total de 100 kg y cada herramienta tiene una masa de 10,0 kg. ¿Con qué velocidad llega el astronauta a la nave?

- a) 1.11 m/s
- b) 1.25 m/s
- c) 2.35 m/s
- d) 4.29 m/s
- e) 2.49 m/s

Fuente: http://www.fing.edu.uy/if/cursos/fis1/parcialesold/2006/p1_06_v1.pdf (Física General 1, Facultad de Ingeniería, UdelaR)

Dos de los **índices** más comúnmente usados para determinar la calidad de la prueba son el *índice de dificultad* y el *índice de discriminación*, dentro de la Teoría Clásica de los Test (TCT), aplicables a cada ítem de la misma.

Índice de dificultad (ID)

El **ID** se define como la relación entre el número de alumnos que responden correctamente el ítem y el número de alumnos que intentaron resolverlo:

ID = A / N donde A es el número de aciertos y N el número total de alumnos.

Su valor varía entre 0 (pregunta muy difícil que ha sido respondida incorrectamente por todos los alumnos o no respondida) y 1 (muy fácil ya que ha sido respondido por todos los alumnos). Es común clasificar los ítems, según su ID, en tramos de respuesta correcta o tramos de dificultad, lo cual facilita el trabajo a los efectos de almacenarlos en un banco de datos para ser re-utilizado en otras pruebas.

A través del I. Disc. de un ítem dado, es posible diferenciar entre los estudiantes que han llegado a dominar un objetivo propuesto (grupo alto) y los estudiantes que no alcanzan a dominarlo (grupo bajo)

$$\text{I. Disc.} = \frac{\text{grupo alto} - \text{grupo bajo}}{N/2}$$

Índice de Discriminación (I. Disc.)

El I. Disc. varía de 0 a 1. Los ítems cuyo índice de discriminación es menor que 0.20 se consideran no discriminantes o no satisfactorios. Los de I. Disc. entre 0,20 y 0,29 se consideran poco satisfactorios y los de más de 0,40 son los ítems con buena discriminación.

Otro índice de discriminación que se puede utilizar es el de *correlación-punto biserial* de los ítems. A los efectos de su cálculo se considera la relación

existente entre el puntaje de cada ítem alcanzado por un estudiante y el puntaje total que dicho estudiante obtiene en la prueba. Así es posible saber en que medida un ítem es capaz de discriminar entre alumnos que tiene un buen desempeño en la prueba y los que no lo tienen.

3.- Corrección y puntuación de las pruebas

En relación a la corrección y puntuación de las pruebas, el valor asignado a cada ítem, debe dar cuenta del aporte del mismo a la prueba total. En el caso de máxima homogeneidad de la prueba, se le aplicará un punto a cada ítem o igual puntuación a cada uno. En este caso la suma de las puntuaciones parciales es la denominada "puntuación directa". No debe dejar de considerarse la aleatoriedad en las respuestas de éste tipo de pruebas, esa probabilidad de acierto es proporcional al número de opciones de respuesta.

Muñiz (1994) plantea una posibilidad de compensación a este problema desde un punto de vista matemático, aplicando a la puntuación directa, un factor de corrección según la siguiente fórmula:

$$\text{Puntuación corregida} = \mathbf{A - E / (n-1)}$$

Donde **A** representa el número de respuestas acertadas (puntuación directa)

E es el número de respuestas erradas

n es el número de alternativas de respuesta en cada ítem.

Para que esta fórmula sea aplicable a la corrección por azar, se deben de cumplir determinados **supuestos**:

- Los aciertos se deben a que los sujetos saben las respuestas o que contestan al azar, acertando.
- Los errores provienen de que los sujetos desconocen la respuesta y contestan al azar, fallando.
- Cuando los sujetos desconocen la respuesta y contestan al azar, todas las alternativas del ítem son equiprobables.

Hay quienes discrepan con la corrección del puntaje aludiendo que el alumno se inhibe al responder si sabe que será penalizado por error, prefiriendo no responder antes que equivocarse.

Para otros, este factor contribuye a una mayor equidad pues para quien no domina la asignatura hay un "plus" por probabilidad que se ve muy disminuído en quien sí conoce los contenidos a evaluar y de entrada descarta ciertos distractores.

En cualquier caso siempre se debe de informar, antes de la aplicación de la prueba, a los estudiantes, si se penalizará el error o no pues esto puede ser determinante en el criterio a seguir en las respuestas.

4.- Presentación de la prueba

En relación a la presentación de la prueba deben considerarse aspectos tales como:

- Instrucciones para su realización
- Formato de las preguntas
- Formato de las respuestas

En las instrucciones se deben especificar aspectos tales como: la *presentación de la pregunta* (ej. elija una de las cuatro opciones presentadas) o *de la respuesta* (ej. marque con una cruz o califique con un número de 1 a 5), *la duración de la aplicación* (ej. Ud. dispondrá de 45 minutos para la respuesta a las siguientes preguntas) o el *rigor en las expresiones* que pautan el tipo de respuesta a dar.

Las preguntas y las respuestas pueden presentarse en un mismo formato o por separado, siendo más sencilla la respuesta en el primer caso.

5.- Críticas a las pruebas objetivas

- Se suele considerar que este tipo de pruebas evalúa esencialmente conocimientos *memorísticos* en desmedro del razonamiento.
- El hecho de elegir la mejor opción en una respuesta refuerza más el *pensamiento selectivo* que los procesos mentales dirigidos a la construcción del conocimiento.
- El *azar* puede resultar un elemento de distorción en la evaluación del conocimiento ya que puede primar en la elección de la respuesta por sobre el razonamiento acerca de la cuestión planteada.
- La brevedad y estereotipificación de las pruebas puede facilitar la copia entre estudiantes.
- La preparación y el diseño es costoso por su propia presentación, su elaboración es compleja ya que exige gran dedicación de recursos humanos calificados así como materiales.

6.- Ventajas de las pruebas objetivas

- Es de fácil aplicación y corrección lo que las torna de elección en contextos de masividad estudiantil.
- Tiene posibilidad de abarcar amplios dominios de aprendizaje.
- Permite detectar errores conceptuales e informar sobre la calidad de los aprendizajes complejos.
- La corrección de la prueba está menos afectada por aspectos subjetivos inherentes al corrector.
- Posibilita la identificación y cuantificación del nivel de cumplimiento de los objetivos propuestos a los alumnos con la posterior posibilidad de replanteo de los mismos o refuerzo de temáticas no consolidadas.
- Admite la aplicación de múltiples análisis estadísticos cuyos resultados posibilitan las diversas tomas de decisiones posteriores.

7.- A modo de resumen: ¿Qué pasos implica el diseño de una prueba objetiva?

- Desarrollar un **marco teórico de referencia** acerca de cómo se concibe la evaluación (si es equivalente a medición, si es parte del proceso de aprendizaje, si apunta a la comprensión, etc.). Asimismo se debe predeterminar el contenido a evaluar, por ej. si se evaluarán metas propuestas, logros a determinado plazo, desarrollo de ciertas competencias, diferentes tipos de contenidos enseñados, entre otras muchas posibilidades.
- Elaborar una *tabla de especificaciones* adecuada a la prueba a construir.
- Realizar una primera aproximación a la prueba a proponer, sin respuestas de alternativa.
- Aplicarla a una muestra de público equivalente al destinatario (primer pilotaje) y analizar los problemas de los respondentes durante la realización de la prueba.
- Analizar las respuestas obtenidas, clasificándolas en: correctas, parcialmente correctas e incorrectas.
- Reestructurar la prueba considerando los tipos de respuesta obtenidas y los procesos desempeñados por los respondentes.
- Plantear las alternativas de respuestas posibles elaborando los distractores a partir de los errores cometidos en la prueba piloto (ver cuadro perfil del ítem).
- Establecer el nivel de dificultad de las respuestas, ej. difícil, medio y fácil a partir del análisis de los índices de dificultad, índice de discriminación, etc de los diferentes ítems (ver definiciones más adelante).

- Solicitar la revisión por parte de otros técnicos de las posibles actividades y alternativas de respuestas planteadas.
- Organizar varias formas de pruebas preliminares usando las actividades diseñadas.
- Validar las actividades planteadas en las pruebas, aplicándolas a un público suficientemente grande así como validarlas pedagógicamente a partir del juicio de expertos en el área (segundo pilotaje).
- Elaborar la versión final de la prueba considerando los ajustes que surjan de la estadística y de los expertos, en caso que se vaya a aplicar masivamente, ej. una prueba nacional de evaluación.
- La puntuación de la prueba debe ser previamente estandarizada para conocer el nivel de competencia de los respondentes en forma estadísticamente significativa.

Antes de sacar conclusiones basándose en los resultados de la prueba es necesario proceder a un **análisis de la calidad** de la misma como instrumento de evaluación. Es necesario juzgarlo en función de la discriminación y la dificultad de los diferentes ítems así como determinar la validez y confiabilidad del mismo. Esto usualmente se realiza con diferentes paquetes estadísticos tales como el ITEMAN, el SPSS, etc.

BIBLIOGRAFIA

- Bloom, B.S., et al; (1975); *Evaluación del aprendizaje*; Troquel; Buenos Aires.
- De Fleitas A., Moreira F. y Soubirón E.; (2002); *Análisis de Evaluación Censal de Aprendizajes en Ciencias de la Naturaleza en los Novenos Años de las Escuelas Rurales en 2002*; Programa de Modernización de la Educación Media y la Formación Docente. (MEMFOD/ANEP).
- De la Orden Hoz A.; (1991); en Lukas J.F.; Santiago K.; (2004); *Evaluación Educativa*; Alianza Editorial; Madrid.
- Mateo, J.; (2000); *La evaluación educativa, su práctica y otras metáforas*; ICE-Horsori; Barcelona.
- Muñiz J.; (1994); *Teoría clásica de los test*; Ediciones Pirámide; Madrid.1
- Lukas J.F.; Santiago K.; (2004); *Evaluación Educativa*; Alianza Editorial; Madrid.
- Scriven; (1967); en Lukas J.F.; Santiago K.; (2004); *Evaluación Educativa*; Alianza Editorial; Madrid.
- UNESCO; (1979); *Evaluation terms: Unesco Glossary*; Unesco; París.

ANEXO

REDACCIÓN DE ÍTEMS DE OPCIÓN MÚLTIPLE

Fuente: <http://www.uam.es/docencia/ace/24a-redaccion.htm#consejos>

1. CONSEJOS GENERALES

- Cuantos más ítems incluyamos, más fiable y válida será la evaluación. Normalmente pueden incluirse hasta 40 preguntas en una hora.
- El número de preguntas para cada parte del temario debe ser proporcional a la importancia asignada en la docencia.
- Hay que revisar los ítems antes de administrarlos. Debemos asegurarnos de que cumplen los requisitos de redacción que más adelante mostramos. Los ítems deben ser claros, comprensibles y adaptados para aquellos estudiantes que los van a responder. Es conveniente también pedir el juicio de otros expertos en la materia para, en un primer momento, eliminar las preguntas que no estén bien construidas.
- Evitar errores gramaticales, abreviaturas, errores en la puntuación y en la ortografía.
- Minimizar el tiempo de lectura. El tiempo de lectura influye sobre el número de ítems que se pueden responder, y por tanto sobre la fiabilidad y sobre la validez del contenido.
- Evitar las preguntas “con trampa”, pues crean serios problemas de validez ya que es imposible determinar qué se mide realmente y dificultan la resolución. Según Haladyna (1994b) existen siete tipos de ítems que se pueden considerar “preguntas trampa”:
 - Los ítems redactados con la intención de confundir a quien lo responde.
 - Los que tienen un contenido irrelevante.
 - Los ítems con opciones de respuesta muy parecidas.
 - Aquellos cuyo enunciado ofrece información irrelevante para su respuesta.
 - Los que tienen varias respuestas correctas.

- Los ítems que, tanto el enunciado como las alternativas de respuesta, contienen información que no se aprendió.
- Aquellos que se formulan de forma ambigua.

2. ¿CÓMO HACER LAS PREGUNTAS?

- Formular el ítem para que evalúe contenidos objetivos y relevantes. El ítem debe estar ligado a lo que se enseña, a los resultados y objetivos que se pretenden en un curso y a lo que se ha explicado en clase.
- Cada ítem debe reflejar un solo contenido (hecho, concepto, principio, procedimiento). Cuando un ítem trata sobre varias cosas a la vez, es difícil saber qué es lo que está mal aprendido.
- Usar un lenguaje adecuado a las personas que van a responder al test.
- Evitar los ítems concatenados: que cada ítem sea independiente. Un ítem no debe dar información para responder correctamente otro ítem.
- Podemos usar un ítem del que sabemos que está bien redactado como plantilla para redactar otro para otra convocatoria (ítems isomorfos).
- Evitar contenidos que sean susceptibles de opinión.
- Debemos procurar que la tarea de resolver un examen no sea un mero ejercicio de reconocimiento. No preguntar exclusivamente por hechos, sino también por contenidos más elaborados (conceptos, principios, etc.).

3. ¿CÓMO CONSTRUIR EL ENUNCIADO?

- No debe existir ninguna pista semántica o gramatical que permita acertar el ítem sin saber la opción correcta.
- Si se usa un formato de frase incompleta, nunca se deben dejar huecos en blanco entre palabras de la frase o al comienzo de la misma.
- Asegurarse de que el enunciado es claro y preciso. El enunciado debe contener la idea central por la que se pregunta y toda la información necesaria para responder.
- Evitar enunciados largos.

- Formular el enunciado preferiblemente en forma afirmativa. El uso de frases negativas tiene efectos contraproducentes en la comprensión y en el rendimiento. No obstante, cuando se usen formas negativas, las partículas de negación deberían ir en negrita o en mayúsculas, para que el lector tenga en cuenta el sentido de la frase y elabore bien la respuesta.

4. ALTERNATIVAS DE RESPUESTA

- Cuando las opciones de respuesta son números, deben aparecer de mayor a menor; si son fechas, deben respetar un orden cronológico. Cuando las alternativas de respuesta son numéricas y se presentan por intervalos, hay que procurar que las alternativas no se solapen.
- Que todas las opciones tengan una longitud parecida y que incluyan un lenguaje técnico sobre lo que se pregunta.
- No abusar de expresiones inclusivas del tipo: "Todas las anteriores son correctas" o "Ninguna de las anteriores es correcta". Estas opciones de respuesta se incluyen muchas veces por la necesidad de completar el número de alternativas cuando escasea la creatividad en la redacción. Este tipo de expresiones favorece el acierto para quienes siguen estrategias picaras de respuesta.
- Evitar el uso de pistas que llevan a elegir la alternativa correcta. En la opción correcta, no se deben incluir expresiones o palabras que aparezcan en el enunciado. Tampoco se deben incluir opciones incorrectas que resulten absurdas y lleven indudablemente a la elección de la alternativa correcta.
- Evitar el uso de palabras como: siempre, nunca, totalmente, absolutamente y completamente. Estas expresiones suelen aparecer en las alternativas incorrectas, pues, como sabemos, muy pocas veces podemos realizar juicios universales. En consecuencia, la presencia de estas palabras ayuda a descartar la opción como correcta.

5. ALTERNATIVA CORRECTA

- Alternar la posición de la respuesta correcta. Si colocamos la respuesta correcta siempre en el mismo lugar, estamos favoreciendo a aquellas personas que responden siguiendo criterios formales.
- La longitud de la opción correcta debe ser similar a la de las incorrectas. En algunos estudios se encuentra que la respuesta correcta se redacta con más precisión y longitud que las incorrectas, lo que facilita el acierto del ítem sin conocer la opción correcta.

6. DISTRACTORES O ALTERNATIVAS INCORRECTAS

- Usar distractores coherentes. Deben ser descartados por aquellos que tienen un alto nivel de conocimiento, y resultar aceptables para aquellos que no saben tanto.
- Para redactar un buen distractor podemos usar los errores más comunes en los estudiantes.
- Usar expresiones técnicas. Los distractores que contiene un lenguaje propio del área de conocimiento que se evalúa, aparecen como respuestas correctas a aquellos que no manejan muy bien el contenido.
- Redactar las opciones mediante contenidos que son verdaderos pero que no responden a lo que se pide.
- Evitar alternativas "graciosas". Una opción formulada con sentido del humor se descarta automáticamente como correcta y convierte al ítem en algo "artificialmente fácil".