

SOFTWARE, INSTRUMENTACIÓN Y METODOLOGÍA

Directrices para la construcción de ítems de elección múltiple

Rafael Moreno, Rafael J. Martínez y José Muñiz*
Universidad de Sevilla y * Universidad de Oviedo

Los ítems de elección múltiple son utilizados en numerosos contextos de evaluación psicológica y educativa. Con demasiada frecuencia las críticas que se hacen a este tipo de ítems tienen más que ver con su construcción defectuosa que con los inconvenientes del formato en sí mismo, que resulta objetivo, fácil y económico de corregir, y muy versátil, adaptándose a todo tipo de contextos y contenidos. El presente artículo destaca la necesidad de contar con un conjunto de directrices eficientes que guíen la construcción de dichos ítems, sustituyendo el modo intuitivo y poco sistemático seguido con frecuencia. Se revisan las directrices existentes y se identifican los diversos problemas que dificultan su aprovechamiento, tales como el elevado número de directrices propuestas, redundancias entre ellas y ciertas imprecisiones. A partir de esa revisión se seleccionan las directrices de contenido relevante y se reorganizan en un conjunto más parsimonioso y estructurado constituido por doce directrices básicas, acompañadas de ejemplos e ilustraciones. Se espera que esta propuesta resulte de utilidad como guía para los profesionales e investigadores que deseen construir ítems de elección múltiple o analizar los ya existentes.

Guidelines for the construction of multiple choice test items. Multiple choice test items are currently used in several areas of psychological and educational assessment. All too often, criticism made of these items concentrates more on their defective construction than the difficulties with the format itself; despite this, the format is objective, easily and cheaply corrected, and its versatility makes it adaptable to many contexts and contents. This paper highlights the need for a set of effective guidelines to assist in the construction of these items, thereby replacing the intuitive and unsystematic approach generally used. Existing guidelines are revised, as are several questions that stop them working properly, such as the excessive number of guidelines, overlapping between them and inaccuracies. Following this revision, the guidelines with important content are selected and reorganized in a more practical and structured set consisting of twelve basic guidelines, accompanied by examples and illustrations. It is hoped that this proposal will be useful as a guide for teachers or researchers who have to build multiple choice items, or analyze already existing ones.

Cuando se lleva a cabo la elaboración de un test, una primera tarea importante es la construcción de los ítems que lo conforman. Sin embargo, si se revisan los manuales de psicometría clásica y los más recientes de Teoría de Respuesta a los Ítems puede observarse que dedican muy poca atención a la tecnología para la construcción de los ítems en comparación con otros tópicos asimismo relevantes, como la fiabilidad del test, la validez o el análisis de los ítems. La explicación a este hecho es la asunción de que la cons-

trucción de los ítems depende más del ingenio e inspiración del constructor que de la aplicación sistemática de una tecnología, confiándolo todo a que los análisis estadísticos a posteriori permitan detectar los ítems inapropiados. Afortunadamente esta filosofía ha ido cambiando y en la actualidad disponemos de trabajos y programas de investigación centrados en la construcción rigurosa de los ítems, bien recogidos en manuales como los de Haladyna (1994) y Osterlind (1998). De acuerdo con Muñiz y García-Mendoza (2002), varias son las razones que han originado este cambio: a) el predominio que han tenido en los últimos años los modelos de Teoría de Respuesta a los Ítems, colocando todo lo relacionado con el ítem como unidad básica de medida en el centro del escenario psicométrico; b) la aparición de los Tests Adaptativos Informatizados (Olea, Ponsoda y Prieto, 1999; Van der Linden y Glass, 2000), que exigen la elaboración continua de ítems para reponer

los Bancos de Ítems utilizados; c) el gran desarrollo de las técnicas para detectar el funcionamiento diferencial de los ítems (Cammill y Shepard, 1994; Fidalgo, 1996; Holland y Wainer, 1993), que obliga a indagar con rigor qué hace que un ítem funcione de diferente modo para distintas poblaciones; d) la interacción entre los modelos psicométricos y la psicología cognitiva (Frederiksen, Mislevy y Bejar, 1993; González-Romá y Espejo, 2003; Prieto y Delgado, 1996, 1999, 2003), que lleva a construir con precisión los distintos componentes de los ítems para poder analizar los procesos psicológicos implicados en su resolución; e) la irrupción de modelos de evaluación alternativos a los convencionales, tales como la denominada evaluación auténtica (Bravo y Fernández del Valle, 2000; Hakel, 1998; Powell, 1990), que ha obligado a mejorar y justificar los ítems de los tests convencionales. Además de estas razones, la combinación del ordenador con otros medios audiovisuales, realidad virtual e internet está abriendo posibilidades insospechadas en la confección de los ítems (Parshall y Balizet, 2001; Parshall, Davey y Pashley, 2000; Zenisky y Sireci, 2003).

Este es el marco general en el que se encuadra nuestro trabajo, cuyo objetivo central es ofrecer a los investigadores y profesionales directrices claras y precisas que les permitan mejorar la construcción de los ítems de elección múltiple, no dejando esta importante labor a la mera intuición e inspiración del momento.

Como es bien sabido, existen diferentes formatos para los ítems. Algunos exigen construir la respuesta, como las *preguntas abiertas*, en las que el sujeto debe elaborar totalmente lo que se le plantea; por ejemplo, «Enumere tres huesos del tórax humano», o «Describa la función clorofílica». También son abiertos los ítems de *completar frases*, que piden al sujeto que añada uno o más elementos de una oración, como por ejemplo, «El ..., la tibia y el ... son ... de las extremidades inferiores humanas».

Otros formatos en cambio ofrecen al sujeto diversas opciones de respuesta, pidiéndole elegir las que reconozca como correctas. El formato más usual es el de *elección múltiple*, consistente en un enunciado o pregunta que se completa con varias opciones de respuesta entre las que el sujeto debe identificar la única correcta. El resto de formatos pueden considerarse variaciones del anterior (ver Tabla 1). El ítem es denominado de *elección alternativa* si presenta sólo dos opciones de respuesta, o de *verdadero-falso* si son esas las opciones ofrecidas. Una versión del anterior es el formato *verdadero-falso múltiple* que pide al sujeto valorar con esas alternativas varios enunciados referidos a un mismo contenido. Otros formatos presentan dos conjuntos de opciones que han de utilizarse de determinados modos: el de *emparejamiento* pide al sujeto asociar los elementos de uno y otro conjunto, y el de *elección múltiple complejo* propone elegir la opción correcta de una serie referida a un conjunto previo de opciones. Por último, ítems de algunos de los formatos de elección múltiple, especialmente el usual, el de elección alternativa y el verdadero-falso, son presentados a veces en grupos referidos a un mismo contexto o referente, dando lugar al formato *conjunto dependiente de un contexto*.

Los ítems de elección múltiple exigen a quien los elabora un mayor esfuerzo y pericia que aquellos más abiertos que sólo plantean la pregunta, aunque hoy por hoy tienen la ventaja de una evaluación más fiable. Por ello, los de elección múltiple son un recurso muy adecuado para estudiar amplias poblaciones de sujetos. A pesar de tales ventajas, estos ítems son vistos con reticencia por cuanto sólo serían adecuados para evaluar tareas sencillas y resultados de procesos más que estos. Aunque tales carencias pueden estar presentes en muchos ítems de este tipo, no son inherentes a ellos; con ítems

de elección múltiple adecuadamente construidos resulta posible evaluar tareas cognitivamente complejas, así como todos los pasos que deseen evaluarse de un proceso, como se pone de manifiesto ampliamente en la práctica de la evaluación (Haladyna, 1994).

Tabla 1
Formatos de ítems de elección múltiple

<p>Elección múltiple usual</p> <p>La fórmula C_3H_{10} corresponde al siguiente hidrocarburo:</p> <ol style="list-style-type: none"> 1. Etano 2. Propano 3. Butano 						
<p>Elección alternativa</p> <p>Una propiedad de las células madres o estaminales es su:</p> <ol style="list-style-type: none"> 1. Unipotencialidad 2. Multipotencialidad 						
<p>Verdadero-Falso</p> <p>La adenina es una de las bases químicas del ADN:</p> <ol style="list-style-type: none"> 1. Verdadero 2. Falso 						
<p>Verdadero-Falso múltiple</p> <p>Señale si es Verdadero (V) o Falso (F) que cada una de las siguientes es característica usual de las plantas monocotiledóneas:</p> <ol style="list-style-type: none"> 1. Son angiospermas 2. Tienen tallo ramificado 3. Tienen flores 						
<p>Emparejamiento</p> <p>Conecte los siguientes países con sus capitales:</p> <table border="0"> <tbody> <tr> <td>1. Mongolia</td> <td>a. Kuala Lumpur</td> </tr> <tr> <td>2. Bangladesh</td> <td>b. Ulan Bator</td> </tr> <tr> <td>3. Malasia</td> <td>c. Dhaka</td> </tr> </tbody> </table>	1. Mongolia	a. Kuala Lumpur	2. Bangladesh	b. Ulan Bator	3. Malasia	c. Dhaka
1. Mongolia	a. Kuala Lumpur					
2. Bangladesh	b. Ulan Bator					
3. Malasia	c. Dhaka					
<p>Elección múltiple complejo</p> <p>Son capitales de provincias españolas:</p> <ol style="list-style-type: none"> a. Oviedo b. Pamplona c. Vigo <ol style="list-style-type: none"> 1. a y b 2. b y c 3. a, b y c 						
<p>Conjunto de ítems dependientes de un contexto</p> <p>«Siento que ha entrado el otoño. La temperatura ha descendido y se respira humedad. Los árboles entrarán en reposo y mostrarán numerosas tonalidades». Las siguientes preguntas sobre sintaxis se refieren al texto indicado.</p> <p>A. ¿Qué parte del sujeto es «temperatura» en la oración en que aparece?</p> <ol style="list-style-type: none"> 1. Núcleo 2. Modificador directo 3. Modificador indirecto <p>B. En el predicado verbal del que forma parte, «numerosas tonalidades» es:</p> <ol style="list-style-type: none"> 1. Objeto directo 2. Objeto indirecto 3. Núcleo <p>C. «Numerosas» es modificador:</p> <ol style="list-style-type: none"> 1. Término 2. Subordinante 3. Directo 						

Las posibilidades de los ítems de elección múltiple se desaprovecharán, sin embargo, en la medida en que la construcción se realice de manera intuitiva. Como se dijo al comienzo, se han ofrecido distintos conjuntos de directrices para aportar una mayor sistemática a esa tarea (Hoepfl, 1994; Marrelli, 1995; Osterlind, 1998; Roid y Haladyna, 1982). Especial mención merece la de Haladyna y Downing (1989), que sintetiza más de cuarenta taxonomías aparecidas en los cincuenta y cuatro años anteriores. Está compuesta por cuarenta y tres directrices organizadas en distintos epígrafes, tal como puede verse en la tabla 2.

En el primer grupo, las directrices *procedimentales* aconsejan un uso correcto de la gramática, una cantidad no excesiva de material a leer en el ítem, la colocación vertical de las opciones y que haya solo una correcta, así como evitar el formato de elección múltiple complejo. Las referidas al *contenido* señalan el nivel y tipo de pensamiento que puede requerir el ítem, mientras que los dos grupos restantes recogen directrices específicas para cada una de las partes del ítem, enunciado y opciones.

Los propios autores señalan problemas en su taxonomía, tales como ambigüedad de algunas directrices, falta de independencia entre otras y ausencia de integraciones que pudieran aportar mayor parsimonia. Por ello, y quizás también por las dificultades que suponía el elevado número de directrices ofrecidas, apareció una nueva versión (Haladyna, Downing y Rodríguez, 2002) compuesta por 31 que puede verse en la tabla 3.

Observada en detalle, la reducción de directrices es menor de lo que parece puesto que una de ellas (la 28) queda especificada en seis diferentes, resultando por tanto un total de 36. Una comparación detallada de los contenidos de las directrices pone de manifiesto al menos los siguientes cambios entre las dos taxonomías de referencia: i) Integración de categorías: las directrices 1, 2, 18 y 19 de Haladyna y Downing (1989) aparecen como componentes de la 9 en Haladyna, Downing y Rodríguez (2002) y las 14 y 16 antiguas pasan a conformar la nueva directriz 3. ii) Disgregación de una en varias: la directriz 33 de la primera versión se diversifica en las variaciones b y f de la 28 nueva. iii) Incorporación de nuevas directrices: aparecen las 28 d y e. iv) Supresión de otras: las números 8, 12, 40, 41 y 42. Otras 27 directrices permanecen idénticas o muy semejantes.

Las variaciones comentadas no resultan suficientes, ya que la nueva versión sigue presentando similares problemas que su predecesora, incumpliendo características exigibles a toda taxonomía o clasificación. Algunas de las directrices no son mutuamente excluyentes: ocurre con la 7 («Evite ítems con trampas») que en su definición incluye explícitamente aspectos de las directrices 2

(«Evite contenidos triviales»), 5 («Evite contenidos muy específicos o muy generales»), 14 («Asegúrese de que el sentido del enunciado resulta muy claro»), 16 («Evite adornar el texto en exceso»)

<i>Tabla 3</i> Directrices para la construcción de ítems de elección múltiple (Haladyna, Downing y Rodríguez, 2002)	
Referidas al contenido	
1.	Cada ítem debería reflejar un contenido específico y una única conducta mental específica, tal como sea requerido en las especificaciones del test (tabla de doble entrada, proyecto del test)
2.	Base cada ítem en un contenido importante para el aprendizaje; evite contenidos triviales
3.	Use material novedoso para evaluar el aprendizaje de alto nivel. Cuando los utilice en un ítem, parafrasee el lenguaje de los libros de texto, o el lenguaje utilizado durante la instrucción, para así evitar evaluar el mero recuerdo
4.	Mantenga el contenido de cada ítem independiente del contenido de otros ítems del test
5.	Al escribir ítems de elección múltiple, evite contenidos muy específicos o muy generales
6.	Evite ítems basados en opiniones
7.	Evite ítems con trampas
8.	Use un vocabulario sencillo para el grupo de estudiantes que están siendo evaluados
Referidas al formato	
9.	Del formato convencional de elección múltiple utilice la interrogación, completar frases, la mejor respuesta, elección alternativa, verdadero-falso, verdadero-falso múltiple, emparejamiento, los conjuntos de ítems y los dependientes de contexto; sin embargo, evite el formato de elección múltiple complejo (el tipo K)
10.	Construya el ítem de forma vertical, no horizontal
Referidas al estilo	
11.	Corrija y pruebe los ítems
12.	Utilice una gramática, puntuación, mayúsculas y minúsculas y deletreo correctos
13.	Mínimice la cantidad de lectura en cada ítem
Redacción del enunciado	
14.	Asegúrese de que el sentido del enunciado resulta muy claro
15.	Incluya la idea central en el enunciado y no en las opciones
16.	Evite adornar el texto en exceso (palabrería excesiva)
17.	Expresé el enunciado de manera afirmativa, evitando términos negativos tales como NO o EXCEPTO. Si usa términos negativos, hágalo con sumo cuidado y asegúrese que aparecen en mayúsculas o negritas
Redacción de las opciones	
18.	Escriba tantas opciones como pueda, si bien la investigación sugiere que con tres es suficiente
19.	Asegúrese que sólo una de esas opciones es la respuesta correcta
20.	Varié la colocación de la respuesta correcta según el número de opciones
21.	Coloque las opciones en un orden lógico o numérico
22.	Construya las opciones independientes entre sí, no deben solaparse
23.	Mantenga a las opciones homogéneas en contenido y estructura gramatical
24.	Escriba las opciones con una longitud aproximadamente igual
25.	La opción <i>Ninguna de las anteriores</i> debe usarse con mucho cuidado
26.	Evite la opción <i>Todas las anteriores</i>
27.	Escriba las opciones de forma afirmativa, evite términos negativos tales como NO
28.	Evite dar pistas sobre la respuesta correcta, tales como: <ol style="list-style-type: none"> a. Determinantes específicos como siempre, nunca, completamente y absolutamente b. Asociaciones por sonido similar y opciones idénticas o parecidas a términos del enunciado c. Inconsistencias gramaticales que indiquen al sujeto la elección correcta d. Opción correcta destacada e. Pares o tríos de opciones que indiquen al sujeto la opción correcta f. Opciones claramente absurdas o ridículas
29.	Haga plausibles todos los distractores
30.	Use errores usuales de los estudiantes para escribir los distractores
31.	Use el humor si es compatible con el profesor y con el ambiente de aprendizaje

<i>Tabla 2</i> Número de directrices por epígrafes en Haladyna y Downing (1989)	
Tipos de directrices	Número
I. Advertencias generales	
A. Procedimentales	7
B. Referidas al Contenido	10
II. Construcción del enunciado	6
III. Desarrollo de las opciones	12
A. Opción correcta	2
B. Distractores	6
TOTAL	43

y 19 («Asegúrese que sólo una de las opciones es la respuesta correcta»). Otras directrices tienen un contenido similar, como la 13 («Minimice la cantidad de lectura en cada ítem») y la ya citada 16. Sin embargo, y a pesar de esta semejanza, ambas están clasificadas en distintos apartados. En sentido similar, la directriz 8 sobre vocabulario simple y adaptado a los sujetos está incluida en el apartado de *contenido del ítem* y no en el de *estilo* como sería de esperar. Por último, las definiciones de algunas directrices son imprecisas, como la 5, referida sin mayor caracterización a ítems «muy específicos», o la 1 que menciona una «única conducta mental específica».

Estos problemas y el aún elevado número de directrices dificultan la utilización de la taxonomía de Haladyna, Downing y Rodríguez (2002). En tal situación, el presente trabajo pretende continuar el avance partiendo de lo ya logrado. Plantea una nueva versión para recoger en lo fundamental el mismo contenido de la taxonomía de referencia pero haciéndolo con mayor parsimonia, para ser más útil a los profesionales que deseen redactar ítems de elección múltiple.

Procedimientos para aumentar la parsimonia

En primer lugar, la nueva versión debe excluir directrices que estén referidas a aspectos colaterales a su objetivo. Por ello no debería incluir la mayor parte de la directriz 9 de Haladyna, Downing y Rodríguez (2002) dedicada a enumerar diferentes formatos de elección múltiple, puesto que las directrices pretendidas han de referirse a los aspectos de construcción comunes a los diferentes formatos y no a estos. Por una razón similar, debe excluir también la referencia a procedimientos auxiliares de la construcción, como hace la directriz 11 de Haladyna, Downing y Rodríguez (2002) «Corrija y pruebe los ítems».

En segundo lugar, deben reorganizarse todas aquellas cuyo contenido relevante no puede ser omitido. Con tal fin resulta conveniente tener en cuenta el sentido u objetivo de los ítems de elección múltiple. Estos se construyen para evaluar algún contenido, como por ejemplo el rendimiento en una materia académica. Por ello debe disponerse de una serie de directrices sobre el contenido a incluir en el ítem y sobre cómo plasmarlo en este. Otro conjunto de directrices deben referirse a las diversas opciones de respuestas al constituir el elemento diferenciador de estos ítems. Se trata de facilitar que el sujeto que conozca la opción correcta pueda identificarla y el que no la conozca no encuentre pistas indebidas sobre ella. Así pues, los tres aspectos referentes citados, contenido a incluir, su expresión en el ítem y las opciones de respuesta, son los necesarios para configurar los apartados de la taxonomía pretendida.

Por último, la reorganización pretendida debe reducir el número de directrices de cada apartado, eliminando redundancias e integrando varias como casos particulares de alguna de las nuevas.

Nuevas directrices

Conforme a los objetivos y procedimientos señalados, se ofrece el siguiente conjunto de doce directrices para la construcción de ítems de elección múltiple, ilustradas con comentarios y ejemplos y sintetizadas en la tabla 4.

A. Elección del contenido que se desea evaluar

1. Los contenidos a recoger en los ítems de una prueba deben ser una muestra representativa de la aptitud, rendimiento o cualquier otro aspecto que se desee evaluar, el cual debe-

ría estar especificado previamente y sin solapamientos entre sus elementos en una tabla lo más exhaustiva posible. En las llamadas tablas de especificación se enumeran los contenidos, incluyendo términos, nociones y competencias considerados relevantes para la temática que interese evaluar. Todos ellos, o al menos una muestra lo más representativa posible, deberían estar presentes en el conjunto de los ítems de una prueba.

Debe evitarse por tanto que el ítem se refiera a contenidos triviales y diferentes al objetivo. Un ejemplo de lo que *no* debe hacerse sería el siguiente ítem (señalado en cursiva como otros inadecuados de más adelante) si fuera utilizado para evaluar conocimientos de Historia, cuando en todo caso lo que evalúa sería memoria fotográfica.

El cronograma sobre la Revolución Francesa del libro de texto aparece aproximadamente en la página:

1.	50
2.	100
3.	150

2. La representatividad respecto a lo que se desea estudiar en una determinada población de sujetos deberá marcar las características de cada ítem en cuanto a lo sencillo o complejo, concreto o abstracto, memorístico o de razonamiento de lo que plantee y en cuanto a la expresión formal o informal, verbal, numérica o gráfica utilizada.

El ajuste respecto a lo que se desea evaluar es el criterio clave para decidir acerca del contenido y forma del ítem. Recuérdese esta clave, que se encontrará aplicada en las directrices específicas que aparecen más abajo. Ninguna decisión es correcta por sí misma, pues dependerá de lo que se desee evaluar, de los sujetos y del medio en el que se desee hacerlo. Por ejemplo, un ítem verbal será adecuado en una prueba de Historia, pero probablemente inadecuado para evaluar buena parte de contenidos algebraicos. Un ítem memorístico será irrelevante para evaluar razonamiento, pero será pertinente si se trata de evaluar memoria. Si por el contrario se pretende evaluar alguna capacidad no memorística, ella será la clase de tarea que debe plantear el ítem, tal como se hace en el siguiente ejemplo.

¿Qué número continúa la siguiente serie?

1, 3, 7, 15, 31, 63...
1. 73
2. 85
3. 127

B. Expresión del contenido en el ítem

3. Lo central del contenido elegido debe expresarse en el enunciado, haciendo de cada opción un complemento que por ello ha de concordar sintáctica y semánticamente con el enunciado.

Desplazar el grueso del contenido a las diferentes opciones desaprovecharía las posibilidades que ofrece el enunciado para dar una mayor amplitud a la expresión; obligaría además a leer una mayor cantidad de texto a veces redundante, lo que puede dificultar una adecuada comprensión del ítem. Ocurriría en el siguiente ejemplo no aconsejable:

En Física, sublimación:

1. *Supone un cambio de materia sólida a materia gaseosa*
2. *Se refiere a un cambio de materia líquida a materia sólida*
3. *Consiste en un cambio de materia gaseosa a materia líquida*

La siguiente versión solucionaría las dificultades anteriores al trasladar la mayor parte del texto al enunciado y evitar repeticiones entre las opciones.

En Física se denomina sublimación a un cambio de materia:

1. Sólida a gaseosa
 2. Líquida a sólida
 3. Gaseosa a líquida
4. La sintaxis o estructura gramatical utilizada debe ser correcta y no oscurecer el contenido que se evalúa; ha de evitarse un ítem demasiado escueto o excesivamente profuso en su redacción, ambiguo o confuso, teniendo especial cautela con las oraciones negativas porque pueden resultar complicadas de entender. Se complicaría indebidamente la comprensión si abusando de las negaciones se escribe por ejemplo un enunciado como el siguiente: «En situaciones difíciles ¿qué no debe dejarse de no hacer?» Salvo que dicho enunciado fuera utilizado para evaluar el dominio de las negaciones por parte de los sujetos, debería ser sustituido por otro más entendible como el siguiente: «En situaciones difíciles ¿qué debería evitarse?»
- En el mismo sentido y salvo que el ítem esté construido para evaluar la capacidad de entender expresiones complejas, la exposición del ítem debe hacerse de la manera más clara posible, para que el sujeto pueda concentrarse en lo que se le pregunta sin encontrar dificultades innecesarias. Ocurre a veces que se complica la redacción en un intento de evitar que el ítem resulte excesivamente sencillo. Ese sin embargo no es el camino. El constructor del ítem debe preguntarse si es relevante plantear un determinado contenido; si lo es, debe incluirlo sin enmascaramiento que impida al sujeto elegir adecuadamente una respuesta. Por ejemplo, *no* debería utilizarse el siguiente ítem para preguntar sencillamente por la capital de Italia y no por el conocimiento de los encubridores utilizados:

La población en la que radican la sede de los órganos políticos, judiciales y legislativos, además de la de muchos agentes económicos, y que constituye la cabeza del Estado llamado usualmente «transalpino» es:

1. *Berna*
 2. *París*
 3. *Roma*
5. La semántica implicada en el vocabulario utilizado debe estar ajustada al contenido y a la población de sujetos evaluados, para que pueda ser comprendida sin dificultad y no desvíe la atención respecto a lo que se pregunta. Esta directriz es paralela a la anterior en lo que se refiere a la semántica. Debe evitarse que el significado de las pala-

bras usadas sea un obstáculo para la comprensión del ítem por el sujeto que lo lee. El siguiente, por ejemplo, sería adecuado en una evaluación del conocimiento sobre el lenguaje usado en el mundo de las drogas, pero no en otro caso.

Un sinónimo de jeringuilla es:

1. Mandanga
2. Pilula
3. Camisa

C. Construcción de las opciones

6. La opción correcta de cada ítem debe ser sólo una, y debe estar acompañada por distractoras que sean plausibles para el sujeto que no conoce la respuesta correcta y fácilmente desechables para el que la conoce.

Un modo de lograr ese tipo de distractoras consiste en utilizar las que sean próximas a la correcta en su significado o en su apariencia.

La radiación alfa consiste en haces de partículas compuestas de:

1. Dos protones y dos neutrones
2. Un protón y dos electrones
3. Dos neutrones y un electrón

Otro modo consiste en utilizar para las distractoras errores que sean comunes en la población en estudio, aprovechando la frecuente e inadecuada asociación con lo que se pregunta; el siguiente ítem usado en evaluación de lengua inglesa utiliza el error frecuente que aparece en la segunda opción.

Indicar qué frase usa INCORRECTAMENTE el adjetivo few o little:

1. It is only a little house
2. I have few interest in economy
4. Few people understand that sentence

7. En un conjunto de ítems, la correcta debe estar repartida entre las distintas ubicaciones posibles para que estas no aporten información indebida.

Al ubicar la correcta de cada uno de los ítems de una prueba, debe evitarse alguna tendencia que pueda ser descubierta por los sujetos, que tendrían así una pista indebida. A veces y sobre todo si la correcta es la primera en ser construida, puede tenderse a dejarla como la primera opción. Para evitarlo, otras veces se extrema la precaución haciendo que no aparezca nunca o casi nunca en esa ubicación. Otras veces se tiende a ubicarla en alguna opción que no sea la primera ni la última. Evítense cualquiera de esas tendencias u otras que distorsionen un reparto equitativo. Este puede hacerse al azar, pero en tal caso ha de cuidarse que la ubicación de las correctas no resulte extraña en el conjunto de las opciones de cada ítem, rompiendo lo expresado en la directriz 10.

8. Las opciones deben ser tres preferiblemente. Si se añade alguna más, debe cuidarse especialmente que también sea plausible la última en ser construida ya que es frecuentemente descuidada.

Añadir más opciones puede ser difícil porque no todos los temas permiten tantas opciones plausibles, por lo cual tiende a construirse otras que el sujeto detecta como falsas con una facilidad indebida. Usar sólo dos es por el contrario más fácil de construir pero aumenta considerablemente la probabilidad de que el sujeto escoja al azar la respuesta correcta.

9. Las opciones deben ser presentadas en vertical para facilitar su lectura

La distribución vertical es un modo de destacar especialmente el ítem y facilitar su lectura así como su diferenciación del resto de los que conforman una prueba. Véase cómo las opciones del siguiente ítem destacan menos que en los expresados hasta aquí de modo vertical:

¿Qué expresión equivale a que alguien no deja hablar en una conversación?

1. No deja meter baza;
2. No deja títere con cabeza;
3. No es baladí

La distribución horizontal de las opciones puede ser adecuada sin embargo en los casos en los que, más que diferenciar las opciones, interese que sean percibidas como una graduación de posibilidades como ocurre en el siguiente ítem.

Indique su grado de acuerdo con utilizar dinero público para proyectos de exploración espacial:

1. Nada o casi nada;
2. Más bien poco;
3. Bastante

10. El conjunto de opciones de cada ítem debe ser organizado u ordenado de modo coherente con el contenido en estudio.

La desorganización del contenido de las opciones en el siguiente ítem obliga al sujeto a hacer un trabajo previo que le distrae del cometido al que debería dedicarse, que no es otro que identificar la respuesta correcta si la conoce.

Son propiedades deseables de los tests:

1. Control, validez y fiabilidad
2. Fiabilidad, eficiencia y validez
3. Eficiencia, objetividad y control

Sería más adecuado evitarle esa tarea previa organizando el ítem como en el siguiente ejemplo:

Son propiedades deseables de los tests:

1. Validez, fiabilidad y control
2. Validez, fiabilidad y eficiencia
3. Objetividad, control y eficiencia

Por la misma razón el siguiente ítem organiza adecuadamente las opciones cuantitativas en un orden ascendente

La expresión algebraica $\{6 \cdot [(34+46)/(12-8)]\}$ es igual a:

1. 80
2. 120
3. 160

11. Las distintas opciones del ítem han de ser autónomas entre sí, sin solaparse y sin referirse unas a otras pues ello introduce dificultades o facilidades indebidas. Al menos por esta razón, deben usarse con mucha cautela las opciones «Todas las anteriores» y «Ninguna de las anteriores».

El siguiente ítem contiene solapamientos entre la primera y tercera opción que deja al sujeto sin poder responder con claridad, puesto que en ambos casos la respuesta sería correcta al ser la católica una de las religiones cristianas.

La religión de Don Juan de Austria era:

1. Católica
2. Protestante
3. Cristiana

12. Ninguna de las opciones debe destacar del resto por ser la única diferente en contenido, en aspectos de apariencia como longitud, estructura gramatical o en algún término que aporte información indebida. Esa homogeneidad puede lograrse haciendo que las opciones sean todas semejantes –en contenido o apariencia– o todas claramente diferentes entre sí como hacen respectivamente los dos ejemplos siguientes.

Las taltuzas son animales:

1. Roedores
2. Anfibios
3. Reptiles

Las taltuzas son:

1. Animales roedores
4. Herramientas de artesanía
5. Canciones étnicas

A veces lo diferencial de una opción la resalta indebidamente como la correcta o como la incorrecta. El primer caso ocurre en el siguiente ítem al ser la correcta la única que concuerda en género gramatical con el artículo con que finaliza el enunciado.

El hueso que va desde la rodilla al tobillo por la parte delantera es la:

1. Tibia
2. Húmero
3. Peroné

El siguiente ítem en cambio destaca indebidamente la segunda opción como incorrecta por ser la única que no parece sonar al idioma francés incluso para el sujeto que no conozca a tal personaje.

¿Quién fue un importante personaje de la Revolución francesa?

1. Baudelaire
2. Washington
3. Robespierre

En general, cuando una de las opciones resulta diferente, puede ocurrir que induzca en el sujeto la duda indebida de si esa diferencia significa que es la correcta o es desechable por incorrecta, lo que supone distraerle de la tarea central de entender el ítem e identificar su opción correcta si la conoce.

Tabla 4

Nuevas directrices para la construcción de ítems de elección múltiple

<p>A. Elección del contenido que se desea evaluar</p> <ol style="list-style-type: none"> 1. Debe ser una muestra representativa del contenido recogido en una tabla de especificación, evitando ítems triviales 2. La representatividad deberá marcar lo sencillo o complejo, concreto o abstracto, memorístico o de razonamiento que deba ser el ítem, así como el modo de expresarlo <p>B. Expresión del contenido en el ítem</p> <ol style="list-style-type: none"> 3. Lo central debe expresarse en el enunciado. Cada opción es un complemento que debe concordar gramaticalmente con el enunciado 4. La sintaxis o estructura gramatical debe ser correcta. Evitar ítems demasiado escuetos o profusos, ambiguos o confusos, cuidando además las expresiones negativas 5. La semántica debe estar ajustada al contenido y a las personas evaluadas <p>C. Construcción de las opciones</p> <ol style="list-style-type: none"> 6. La opción correcta debe ser sólo una, acompañada por distractoras plausibles 7. La opción correcta debe estar repartida entre las distintas ubicaciones 8. Las opciones deben ser preferiblemente tres 9. Las opciones deben presentarse usualmente en vertical 10. El conjunto de opciones de cada ítem debe aparecer estructurado 11. Las opciones deben ser autónomas entre sí, sin solaparse ni referirse unas a otras. Por ello, deben evitarse las opciones «Todas las anteriores» y «Ninguna de las anteriores» 12. Ninguna opción debe destacar del resto ni en contenido ni en apariencia
--

Discusión y conclusiones

Tras la revisión realizada, se proponen doce directrices básicas a seguir para la construcción de los ítems de elección múltiple. El conjunto propuesto supone una reducción del número de directrices respecto a la versión tomada como referencia. Una reducción que sin embargo no excluye ninguna de las relevantes contenidas en dicha referencia. Las directrices 2, 4 y 28 f de la taxonomía de Haladyna, Downing y Rodríguez (2002) quedan incluidas en la 1 propuesta aquí; las 3, 5, 6 y 31 del 2002 quedan entendidas como casos particulares de la directriz 2 arriba presentada; la 15 previa equivale a la 3 nueva, que incluye además la concordancia no in-

cluida en la precedente; las 12, 13, 16, 17 y 27 quedan subsumidas en la nueva directriz 4; las 8 y 14 corresponden a la 5 nueva; las 19, 29, 30 integran la 6 aquí propuesta; la 20 corresponde ahora a la 7; la 18 queda reformulada en la 8 nueva que además añade un aviso sobre la última en ser construida, que no estaba presente en la taxonomía de referencia; la 10 de referencia es expresada en la 9 propuesta aquí; la 21 pasa a la 10; las 22, 25 y 26 integran la nueva 11; por último, las 23, 24, 28 a, b, c, d, e y f son integradas en la 12 nueva. Asimismo, las directrices propuestas no incluyen algunas de la taxonomía de referencia: la 9 y la 11 por la razón expuesta de que se refieren respectivamente a tipos de formatos de elección múltiple y a procedimientos auxiliares de la construcción y por tanto no a los aspectos que aquí interesa obtener; y las 1 y 7 por ser poco precisas y difíciles de entender y además estar ya recogido su sentido en la 4 nueva.

La reorganización de directrices llevada a cabo puede constituir una herramienta útil para el profesional que desee construir ítems de elección múltiple o tenga que analizar los ya existentes. Facilitará además la investigación empírica sobre los distintos aspectos de los ítems de elección múltiple, algo muy conveniente dada la insuficiencia de datos existentes para muchas de las directrices, apoyadas en su defecto en la práctica acumulada con la construcción de ítems (Haladyna y Downing, 1989; Haladyna, Downing y Rodríguez, 2002). Las directrices también pueden resultar de gran utilidad a la hora de estudiar de forma rigurosa la validez de contenido de una prueba, pues no conviene olvidar que en el complejo proceso de validación de los tests el análisis de los contenidos constituye una fase esencial (Elosúa, 2003; Muñiz, 2004).

Como ya hemos señalado previamente, la tecnología derivada del ordenador, combinada con el uso de internet, la realidad virtual, audio, video, etc., está abriendo la posibilidad de construir ítems sofisticados técnicamente (Parshall y Balizet, 2001; Parshall, Davey y Pashley, 2000; Zenisky y Sireci, 2003), que conserven y potencien la objetividad, economía y versatilidad del formato de elección múltiple, razones poderosas para mantener este formato entre nosotros por muchos años. Su peor enemigo es su construcción defectuosa, como se ha visto a lo largo del presente trabajo. Esperemos que las doce directrices propuestas contribuyan a paliar este problema.

Referencias

- Bravo, A. y Fernández del Valle, J. (2000). La evaluación convencional frente a los nuevos modelos de evaluación auténtica. *Psicothema*, 12 (Supl.), 95-99.
- Camilli, G. y Shepard, L.A. (1994). *Methods for identifying biased test items*. London: Sage.
- Elosúa, P. (2003). Sobre la validez de los tests. *Psicothema*, 15, 315-321.
- Fidalgo, A. (1996). Funcionamiento diferencial de los ítems. En J. Muñiz (Ed.), *Psicometría*. Madrid: Universitat.
- Frederiksen, N., Mislevy, R. y Bejar, I. (1993). *Test theory for a new generation of tests*. Hillsdale, NJ: LEA.
- González-Romá, V. y Espejo, B. (2003). Testing the middle response categories not sure, in between, and ? in polytomous items. *Psicothema*, 15, 278-284.
- Hakel, M.D. (Ed.) (1998). *Beyond multiple choice: Evaluating alternatives to traditional testing for selection*. Mahwah, NJ: LEA.
- Haladyna, T.M. (1994). *Developing and validating multiple-choice test items*. Hillsdale, NJ: LEA.
- Haladyna, T.M. y Downing, S.M. (1989). The validity of a taxonomy of multiple-choice test item. *Applied Measurement in Education*, 1 (1), 51-78.
- Haladyna, T.M., Downing, S.M. y Rodríguez, M.C. (2002). A review of multiple-choice item-writing guidelines. *Applied Measurement in Education*, 15 (3), 309-334.
- Hoepfl, M.C. (1994). Developing and evaluating multiple choice tests. *Technology Teacher*, 53 (7), 25-26.
- Holland, P.W. y Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: LEA.
- Marrelli, A.F. (1995). Writing multiple-choice test items. *Performance and Instruction*, 34 (8), 24-29.
- Muñiz, J. (2004). La validación de los tests. *Metodología de las Ciencias del Comportamiento*, en prensa.
- Muñiz, J. y García-Mendoza, A. (2002). La construcción de ítems de elección múltiple. *Metodología de las Ciencias del Comportamiento*, Monográfico, 416-422.
- Olea, J., Ponsoda, V. y Prieto, G. (1999). *Tests informatizados. Fundamentos y aplicaciones*. Madrid: Pirámide.
- Osterlind, S.J. (1998). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats* (2nd ed.). Boston: Kluwer Academic Publishers.

- Parshall, C.G. y Balizet, S. (2001). Audio computer-based tests: An initial framework for the use of sound in computerized tests. *Educational Measurement: Issues and Practice*, 20(2), 5-15.
- Parshall, C.G., Davey, T. y Pashley, P. (2000). Innovative item types for computerized testing. En W.J. van der Linden y C. Glass (Eds.), *Computer-adaptive testing: Theory and practice*. Boston: Kluwer Academic Publishers
- Powell, M. (1990). *Performance assessment: Panacea or Pandora's box?* Rockville, MD: Montgomery County Public Schools.
- Prieto, G. y Delgado, A.R. (1996). Construcción de ítems. En J. Muñiz (Ed.), *Psicometría*. Madrid: Universitas.
- Prieto, G. y Delgado, A.R. (1999). Medición cognitiva de las aptitudes. En J. Olea, V. Ponsoda y G. Prieto (Eds.), *Tests informatizados. Fundamentos y aplicaciones*. Madrid: Pirámide.
- Prieto, G. y Delgado, A.R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15, 94-100.
- Roid, G.H. y Haladyna, T.M. (1982). *A technology for test-item writing*. New York: Academic Press.
- Van der Linden, W.J. y Glass, C. (Eds.) (2000). *Computer-adaptive testing: Theory and practice*. Boston: Kluwer Academic Publishers.
- Zenisky, A.L. y Sireci, S.G. (2003). Technological innovations in large-scale assessment. *Applied Measurement in Education*, 15, 337-362.