



Dirección General de Servicio Civil

ÁREA DE RECLUTAMIENTO Y SELECCIÓN DE PERSONAL

**Guía Orientativa
ARSP-USEEI-003-2022
Diseño y creación de ítems para la
evaluación psicométrica y de conocimientos**

San José, Costa Rica
Setiembre, 2022

**Guía Orientativa
 ARSP-USEEI-003-2022**

Nombre de la Guía: Diseño y creación de ítems para la evaluación psicométrica y de conocimientos		
Elaborado por: Jose Laurian Ramírez Díaz	Cargo: Profesional Analista	Fecha de elaboración: Octubre, 2021
Revisado por: Jéssica Calderón Morera Angie Badilla Araya María Antonieta Chavarría Fernández Kattia Jiménez Porras Martha Sánchez Li Viviana Zamora Beita Patricia Zúñiga Hernández	Cargo: Profesionales Analistas	Fecha de revisión: Octubre, 2021

Aprobado por:	Cargo:	Fecha de aprobación:
David Campos Calderón	Director Área de Reclutamiento y Selección de Personal	Agosto, 2022
Karla Carranza López	Jefatura de la Unidad de Sistematización y Estandarización para la Evaluación de la Idoneidad	Agosto, 2022

Tabla de contenido

Tabla de contenido	4
Introducción	5
Tipos de ítems	6
Generalidades del proceso de construcción de ítems	6
A. Ítems de ejecución máxima	7
B. Ítems de ejecución típica: el uso de los escalamientos de Likert	20
Errores en la creación de ítems	25
Conclusiones	37
Bibliografía	38

Introducción

La creación de ítems de pruebas que se aplican en los procesos de selección de personas es una tarea que conlleva mucho cuidado, ya que se trata del elemento central de las mismas. Es por esto que se hace necesario contar con el bagaje suficiente, que permita que dicho proceso (creación de pruebas de selección) transcurra de la mejor forma posible.

Este documento sistematiza lo anterior con base en dos líneas de análisis: primero, se exploran y explican los formatos más utilizados para construir ítems de respuesta correcta e incorrecta, así como ítems basados en escalamientos de Likert y, segundo, se hace un recuento sobre los errores más comunes que se cometen al construirlos. Asimismo, se ofrecen ejemplos prácticos para ambos casos.

Con lo anterior se espera que este documento ofrezca una ayuda práctica que facilite la creación de pruebas para los procesos de reclutamiento de la Dirección General de Servicio Civil, así como para otras instituciones que requieran invertir recursos en dichos procesos.

Tipos de ítems

Generalidades del proceso de construcción de ítems

Para la construcción de ítems es importante considerar algunos ejemplos dados por Muñiz, Fidalgo, García, Martínez y Moreno (2005) y Haladyna y Rodríguez (2013), tanto en el diseño correcto como incorrecto de los mismos.

Es importante considerar la distinción que hacen Muñiz et al. (2005) sobre dos tipologías generales de ítems: **los ítems de ejecución máxima**, que se utilizan en los llamados tests de poder, cuyo objetivo es el de identificar a personas que poseen habilidades o características “deseables” respecto a quienes no y que se constituyen con base en opciones de respuesta correctas e incorrectas, así como **los ítems de ejecución típica**, utilizados para construir los llamados tests de dominio, los cuales evalúan con base en niveles o puntajes, a partir de una escala particular.

Con base en lo anterior, es preciso aclarar que este documento se dividirá en dos tipos de ítems, según las características respectivas de las respuestas, a saber:

- Ítems de ejecución máxima, los cuales presentan una evaluación basada en respuestas correctas e incorrectas y su uso es recomendable en pruebas que midan conocimiento.
- Ítems de ejecución típica, específicamente aquellos diseñados para ser evaluados mediante escalas de Likert y se recomienda utilizarlos en pruebas que evalúen aspectos que no puedan medirse en parámetros de correcto o incorrecto, por ejemplo como sucede con las competencias.

A. Ítems de ejecución máxima

Para iniciar, es importante mencionar dos aspectos teóricos que deben considerarse al momento de construir ítems con respuestas correctas e incorrectas:

- Es recomendable que estos ítems presenten tres (3) alternativas de respuesta (una correcta y dos distractores plausibles que demanden a la persona un análisis para descartarlas), y que se apeguen al tema o contenidos del encabezado del ítem.
- Se recomienda que las opciones de respuesta se coloquen de forma vertical, no horizontal (Haladyna y Rodríguez, 2013).

No obstante esas recomendaciones, no se descarta el uso de ítems dicotómicos (con dos opciones de respuesta), o con más de tres alternativas de respuesta.

A continuación se presentan los formatos más utilizados para este tipo de ítems, según lo plantean Haladyna y Rodríguez (2013), así como sus ejemplos respectivos.

1. Ítems basados en la comprensión de líneas de conocimiento

Este tipo de ítem requiere que la persona analice sobre un tema o contenido para contestar adecuadamente. Es necesario que el enunciado contenga toda la información requerida para responder, utilizando una redacción sencilla en el

encabezado y en las alternativas de respuesta. A continuación, se presenta un ejemplo:

El estudiante mostró un comportamiento que distrajo al grupo durante el examen que se realizó en clase. ¿Cuál es el significado de “comportamiento que distrajo al grupo” en la oración anterior?

- a) Conducta inmovilizada.
- b) Conducta tranquila.
- c) Conducta escandalosa.

En el ejemplo anterior, la persona debe analizar el concepto de “comportamiento que distrajo al grupo” con base en los significados o definiciones personales que tiene al respecto para seleccionar la respuesta correcta.

Para este caso en particular, la respuesta correcta se justifica en que el “comportamiento que distrajo al grupo” puede entenderse como una conducta escandalosa (hacer ruido) a partir de las consecuencias que tiene para las otras personas, tal como la distracción que les puede generar en un contexto de un examen aplicado en el salón de clase.

2. Ítems diseñados para medir habilidades técnicas

Este tipo de ítem evalúa criterios relacionados con habilidades necesarias para realizar tareas o funciones adecuadamente. De esta forma, la evaluación se relaciona a conocimientos generales (saberes esperados), habilidades lingüísticas, entre otras, tal como se muestra a continuación:

¿Cuál de las siguientes opciones está escrita correctamente?

- a) Sureña persona
- b) Sureña; persona
- c) Persona sureña

En el caso del ejemplo anterior, la persona debe demostrar sus habilidades de escritura, de modo que la opción “c” es la correcta, porque en el idioma español el sujeto siempre se ubica antes del adjetivo, lo cual no sucede con la opción “a” y “b”, unido que para el caso de la opción “b”, se hace un mal uso de la puntuación.

3. Ítems de aplicación de conocimiento y habilidades para seleccionar la respuesta correcta

La diferencia de este ítem respecto al anterior es que, para este caso en particular, se incorporan más de un aspecto relacionado a distintas habilidades para contestar, de modo que este tipo de ítem incorpora varios aspectos, mientras que el anterior es más concreto, tal como se puede apreciar a continuación:

Una entrada al cine tiene un costo de ₡3 700. Alejandra tiene ₡1 400 en su cartera y en su escritorio tiene cuatro monedas de ₡100 y seis monedas de ₡25. En su abrigo encontró una moneda y un billete que suman ₡1 500 (el billete es de color rojo y es el de más baja nominación). ¿Cuánto dinero le falta para ir al cine?

- a) ₡100
- b) ₡250
- c) ₡775

Para este caso, es importante distinguir los elementos a considerar por la persona para responder adecuadamente:

- Debe emplear habilidades aritméticas (sumar y multiplicar).
- Debe construir imágenes mentales a partir de las características de un objeto, específicamente del color y valor del billete que, para el caso específico del Colón costarricense, tiene la siguiente apariencia:



Conociendo lo anterior, la respuesta correcta para este ejemplo es la “b”:
₡250.

4. Ítems que presentan una afirmación, oración o premisa parcial, que debe completarse con la respuesta correcta

Este tipo de ítem demanda una comprensión conceptual sobre un tema en donde, para contestar correctamente, la persona debe tener dominio técnico o teórico de la temática.

Un aspecto que debe cuidarse con el uso de este tipo de ítems es que la evaluación no se base únicamente en criterios memorísticos.

A continuación, se ofrece un ejemplo del mismo:

El proceso de fusión sucede cuando las moléculas

- a) dejan de moverse.
- b) se mueven de un sitio a otro.
- c) penetran en una célula.

Como puede observarse, la afirmación o encabezado del ítem se basa en un planteamiento que debe completarse con una respuesta correcta, la cual corresponde a la opción “c”, lo cual se basa en temas que nunca cambiarán su significado o contexto, es decir, la respuesta correcta del ítem siempre será la misma.

Otra consideración con respecto a este ejemplo es la redacción, en la que se resaltan dos aspectos:

- El encabezado del ítem no lleva ninguna puntuación. Ocasionalmente se pueden encontrar ítems que concluyen la afirmación con dos puntos, siendo lo más recomendable que se omita este tipo de prácticas para ofrecer una lectura más fluida.
- Las opciones de respuesta inician en minúscula, lo cual es recomendable para que la persona relacione el encabezado del ítem con cada una de estas alternativas, siendo que ambos elementos se complementan.

5. Ítems basados en un formato de mejor respuesta (todas las opciones son correctas pero una alternativa es la mejor)

En estos ítems todas las opciones de respuesta son correctas, pero una de ellas es la mejor o la más adecuada. A continuación, se presenta un ejemplo:

¿Cuál es el dispositivo de seguridad con mayor efectividad para prevenir lesiones producidas por accidentes automovilísticos, ubicado en la parte delantera de la carrocería del vehículo?

- a) Luces intermitentes o de parqueo
- b) Parachoques delantero
- c) Sensor de distancia delantero

Para responder el ítem la persona debe considerar dos aspectos clave del encabezado: primero, la cualidad preventiva del dispositivo al que se hace

referencia y, segundo, su ubicación (“...en la parte delantera de la carrocería del vehículo”).

La opción “a” refiere una función de las luces de dirección que, generalmente, se utiliza para indicar que el vehículo está parqueado. Asimismo, dichas luces se ubican adelante y atrás del vehículo, de modo que no es la respuesta más adecuada si se relaciona al encabezado.

Con respecto a la opción “b”, aunque el parachoques es el delantero, el mismo no previene accidentes en sí, sino que su propósito es minimizar el impacto del choque para las personas que viajan en la cabina.

La opción “c” hace referencia a un sensor delantero de distancia y es, por tanto, la alternativa más adecuada, ya que su propósito es totalmente preventivo (advierte si la distancia es corta) y se encuentra en la parte delantera del automóvil, por lo que se cumplen las premisas del encabezado y, por lo tanto, es la alternativa más adecuada.

6. Ítems de espacios en blanco

Este tipo de ítem refiere un dominio conceptual de temas particulares en contextos específicos. Se trata de que la persona seleccione elementos, conceptos o aspectos que son los más adecuados en situaciones específicas.

A continuación, se presenta un ejemplo de lo anterior:

Negar el reconocimiento de las necesidades de minorías es un tipo de violencia _____, mientras que los insultos repetidos hacia alguien por su aspecto físico es un ejemplo de violencia _____.

- a) Estructural; emocional
- b) Agresiva; estructural
- c) Emocional; psicológica

Para el diseño y construcción de este tipo de ítem se debe tener cuidado de algunos aspectos tales como:

- El encabezado del ítem debe explicar las condiciones necesarias para que las personas comprendan el planteamiento del mismo.
- Las respuestas deben referirse a los conceptos clave para responder adecuadamente.

Para el caso de este ejemplo, nótese que en ambas ocasiones se deja claro que los conceptos que deben seleccionarse en la respuesta corresponden a tipos de violencia. Asimismo, para cada caso se ofrece un contexto general, de modo que la persona contará con los insumos para responder con base en ello, siendo la opción correcta la opción “a” para este caso.

Es importante recalcar que la literatura recomienda un uso limitado o discreto de este tipo de ítem, ya que se basa en instrucciones que pueden mal interpretarse por las personas, dependiendo de condiciones tales como el lugar de residencia, el género, nivel educativo, entre otros.

7. Ítems sin cifrar o de opción múltiple

Estos ítems presentan más de una alternativa de respuesta correcta, de modo que su valor o puntaje corresponde al promedio de los aciertos que tenga la persona. Es importante indicar que, por su naturaleza, no existe una recomendación específica sobre el número idóneo de las opciones de respuesta que deben presentarse, y suele utilizarse una lista de palabras o números para indicarlas.

Este tipo de ítem se utiliza para evaluar aspectos relacionados a un tema en particular, tal como se muestra a continuación:

Seleccione cuatro valores que sean múltiplos de 30

- a) 2
- b) 3
- c) 5
- d) 6
- e) 8
- f) 9
- g) 10
- h) 12
- i) 13

En el ejemplo anterior son cinco las respuestas correctas: “a”, “b”, “c”, “d” y “g”, ya que 2, 3, 5, 6 y 10 son múltiplos de 30, siendo que el residuo de su división es igual a cero.

8. Ítems de formato ambos / ninguno (presentan cuatro alternativas de respuesta, pero solo dos son correctas, ya sean verdaderas o falsas)

Este tipo de ítem presenta una estructura en la que la respuesta correcta es solo una, pero se deben analizar todas las opciones para determinar cuál es la verdadera, incluyendo siempre una alternativa que indica que ninguna de las demás opciones es correcta, por lo que puede tener más de tres (3) opciones de respuesta, tal como se muestra en el siguiente ejemplo:

¿Cuál de los siguientes factores se reconoce como una causa del calentamiento global?

- a) Disminución del uso de combustibles fósiles
- b) Reforestación
- c) Ambas opciones (a y b)
- d) Ninguna de las opciones (a o b)

Para el caso del ejemplo anterior, la opción correcta es la “d”, siendo que las alternativas “a” y “b” no son causas del calentamiento global.

9. Ítems dicotómicos de opción alternativa

Estos ítems presentan dos opciones de respuesta en donde una es correcta y la otra incorrecta, por tanto, presentan con una probabilidad de acierto es del 50%.

Una variante de este formato presenta ambas alternativas de respuesta correctas, de modo que la persona perdería la mitad del puntaje por marcar solo una, lo que debe indicarse claramente en las instrucciones generales o en el encabezado del ítem. Asimismo, para este formato no debe disponerse de dos respuestas incorrectas, ya que esto sería un distractor desventajoso.

A continuación, se presenta un ejemplo:

¿Cuál es un ejemplo de metáfora?

- a) Él camina rápidamente todo el tiempo, casi corriendo
- b) Sus ojos son tan hermosos como el brillo de dos luceros

Para el caso del ejemplo anterior, la respuesta correcta es la opción “b”.

10. Ítems dicotómicos de Verdadero-Falso / Sí-No

Este tipo de ítem siempre representará un 50% de posibilidades de acertar la respuesta correcta.

En el ejemplo a continuación se dispone de una tipología de Verdadero-Falso, aunque este formato puede presentar distintos valores tales como Sí-No, Siempre-Nunca, Correcto-Incorrecto, entre otros. Su uso es muy efectivo para situaciones o afirmaciones que representen extremos y pueden disponerse individualmente o en secuencias.

En el paréntesis, escriba una “V” si la premisa es verdadera o “F” si es falsa

- a) El sol es una estrella ()
- b) Marte y Júpiter tienen dimensiones iguales ()
- c) El agua y el alcohol hierven a 100° C ()
- d) El sonido viaja más rápido que la luz ()
- e) El agua es un recurso agotable ()
- f) La moneda costarricense es el Colón ()

En el ejemplo anterior, el cual presenta una secuencia de ítems, las afirmaciones verdaderas corresponden a las opciones “a”, “e” y “f”.

11. Ítems de pareo o asociación

Este formato de ítems presenta asociaciones de conceptos y significados relacionados entre sí. En esta estructura debe tenerse el cuidado de ubicar los conceptos en una columna y los significados en la otra, sin mezclarlos, ya que podría generar confusión y distraer innecesariamente.

Para ejemplificar lo anterior se disponen los significados en la columna izquierda y los conceptos en la derecha:

Asocie las premisas de la columna izquierda con las opciones de la columna derecha, según corresponda, indicando la letra de la columna izquierda en el paréntesis de la derecha.

- | | | |
|--|-----|--------------|
| a) Estudia las probabilidades de que sucedan las cosas | () | Psicología |
| b) Estudia el movimiento de objetos como los astros | () | Meteorología |
| c) Estudia los océanos y mares | () | Economía |
| d) Analiza las propiedades y comportamiento del clima | () | Sociología |
| e) Estudia las bases y causas del comportamiento humano | () | Oceanografía |
| f) Estudia el comportamiento sistémico de las sociedades | () | Física |
| g) Analiza el comportamiento del mercado y la economía | () | Estadística |

La secuencia para resolver el ejemplo anterior es “e”, “d”, “g”, “f”, “c”, “b” y “a”.

B. Ítems de ejecución típica: el uso de los escalamientos de Likert

Tal como se mencionó, los escalamientos de Likert corresponden a ítems de ejecución típica¹ que presentan afirmaciones, posturas o juicios, ante los cuales se solicita a las personas que manifiesten su opinión eligiendo alguno de los puntos o categorías predefinidas, asignándoles un valor numérico a cada una (Hernández, Fernández y Baptista, 2014).

Aunque no existen reglas definitivas para su redacción, Morales, Urosa y Blanco (2003) plantean algunas consideraciones:

- a. Los ítems suelen formularse como opiniones con las que se puede estar o no de acuerdo, entendiéndolas como actitudes verbalizadas. También formulan conductas o rasgos personales.
- b. Deben tener una relación bastante clara con el aspecto, dimensión o constructo que se desea medir.
- c. Es necesario que la redacción sea clara para que sean comprendidos por la mayoría de personas, evitando el uso de expresiones negativas (palabras como “no” , “nunca” o similares), e incluir una sola idea en cada ítem (unidimensionalidad).
- d. Deben ser discriminantes, es decir, que algunas personas estén de acuerdo con ellos y otras no.
- e. Es necesario formular ítems repetitivos, es decir, que se exprese una misma idea de modo distinto entre varias afirmaciones, con el propósito de medir adecuadamente el constructo, dimensión o aspecto evaluado.

¹ Las respuestas muestran cuál es la postura de la persona que responde.

- f. La redacción preliminar de un ítem debe revisarse por varias personas, con el fin de validar la forma en cómo se plantea la afirmación o idea².

Con respecto al total de valores asignados para recopilar la reacción de las personas, la literatura recomienda entre cinco a siete (Muñiz, Fidalgo, García, Martínez y Moreno, 2005; Hernández, Fernández y Baptista, 2014), aunque existen escalas que funcionan con tres o cuatro inclusive.

Un ejemplo del uso de ítems con base en escalamientos de Likert es el siguiente, el cual se relaciona al costo de los pasajes de avión y los elementos que pueden influir en ello:

Instrucciones: A continuación, se presenta una serie de afirmaciones que debe contestar de la forma más sincera posible, con base en una escala de 1 a 5, según los siguientes criterios:

- 1 = Totalmente en desacuerdo
- 2 = Bastante en desacuerdo
- 3 = Indiferente
- 4 = Bastante de acuerdo
- 5 = Totalmente de acuerdo

No hay respuestas buenas ni malas.

	Afirmación	1	2	3	4	5
1	Viajar en primera clase es costoso debido a la calidad del servicio					
2	La comodidad está relaciona al precio del boleto					
3	El precio de los combustibles influye en el valor de los boletos					
4	El nivel de comodidad durante el vuelo aumenta el precio del boleto					
5	Las aerolíneas definen los precios de forma independiente					
6	El costo de un boleto se relaciona al renombre de la aerolínea					

² La [Entrevista Cognitiva](#) constituye una técnica adecuada para validar preliminarmente la comprensibilidad y claridad de los ítems entre varias personas.

Como puede verse, los ítems con escalamiento de Likert se apegan a las mismas recomendaciones para el diseño de ítems de ejecución máxima, ya que en su conjunto miden un mismo rasgo latente, aunque al medirse varios constructos o dimensiones en una prueba, estos deben disponerse de forma aleatoria, siempre y cuando la escala utilizada sea la misma.

Asimismo, los ítems con escalamiento de Likert pueden redactarse de dos formas:

- Ítems positivos, los cuales presentan situaciones que plantean, como puntaje máximo, el nivel extremo positivo de la escala (Ej. Totalmente de acuerdo, Siempre, Totalmente satisfactorio, entre otros).
- Ítems negativos, que presentan situaciones que plantean, como puntaje máximo, el nivel extremo negativo de la escala y por tanto deben evaluarse de forma inversa (Ej. Totalmente en desacuerdo, Nunca, Totalmente insatisfactorio, entre otros).

Un ejemplo de ítems positivos y negativos pueden ser las siguientes afirmaciones: “Estudiar es divertido” (actitud favorable) y “Estudiar es aburrido” (actitud desfavorable) respectivamente. La persona que conteste de forma sincera lo hará positivamente con uno y de forma contraria con el otro.

A continuación, se presenta un ejemplo más amplio de lo anterior, con base en una escala de cuatro acuerdos, en donde los primeros tres ítems son positivos y los restantes negativos:

Instrucciones: A continuación, se presenta una serie de afirmaciones que debe contestar de la forma más sincera posible, con base en una escala de 1 a 4, según los siguientes criterios:

- 1 = Totalmente en desacuerdo
- 2 = Algo en desacuerdo
- 3 = Algo de acuerdo
- 4 = Totalmente de acuerdo

No hay respuestas buenas ni malas.

	Afirmación	1	2	3	4
1	El liderazgo es una habilidad que aumenta la confianza en el equipo de trabajo				
2	La eficacia de un equipo de trabajo está influida por la gestión de liderazgo				
3	La mejora continua requiere de un liderazgo que motive a las personas a esforzarse				
4	Una característica de un liderazgo activo es pasar por alto las jerarquías sociales				
5	La persona líder cumple solo las normas sociales que le convienen a nivel personal				
6	El liderazgo se basa en manipular a otras personas, sin importar las repercusiones				

Existen otros aspectos de forma que deben tomarse en cuenta a la hora de diseñar y construir ítems de este tipo:

- **Transparencia.** Deben evitarse afirmaciones fáciles de interpretar que generen respuestas idénticas en la mayoría de las personas.
- **Aquiescencia.** A veces las personas tienden a mostrarse de acuerdo con la mayoría de las afirmaciones, aun cuando la respuesta esperada para algunas sea contraria. La inclusión de ítems positivos y negativos puede contribuir a disminuirlo.

- Deseabilidad social. Es común que las personas quieran mostrar su mejor versión en los tests, de modo que tratarán de responder a conveniencia para lograr una visión favorable, manipulando su imagen o, por el contrario, señalando o describiendo posiciones sesgadas sobre los aspectos evaluados en las pruebas (Morales, Urosa y Blanco, 2003; Enríquez y Domínguez, 2010).

De esta forma, los ítems con escalamiento de Likert se constituyen en un formato muy adecuado para pruebas de competencias y similares, por lo que es importante que se consideren los aspectos mencionados en el diseño y construcción de pruebas que evalúan estos aspectos.

Seguidamente, es importante analizar algunas prácticas que se constituyen como errores al momento de crear ítems.

Errores en la creación de ítems

Algunas prácticas pueden repercutir en la fiabilidad de los ítems para recopilar la información requerida, lo cual puede causar confusión y repercutir en la validez³ y confiabilidad⁴ de las pruebas. De esta forma, se presentan algunas situaciones que deben evitarse para prevenir lo anterior.

Error 1: Incluir varios elementos o criterios en un mismo ítem

Los ítems deben evaluar un solo elemento, rasgo o dimensión a la vez, ya que incluir varios aspectos produce ambigüedades que conducen hacia respuestas erróneas, o que generan confusión sobre qué es exactamente lo que se pregunta.

La práctica más común en este tipo de error consiste en utilizar conjunciones como “o” e “y”, tal como se muestra a continuación:

Las moléculas compuestas por tres átomos de oxígeno y la roca fundida o magma son elementos presentes en:

- a) El núcleo y la atmósfera terrestre
- b) Las masas oceánicas y placas continentales
- c) Las capas superiores de la atmósfera y el manto terrestre

Como puede verse, en el encabezado se insinúan elementos relacionados a dos elementos distintos: la atmósfera (oxígeno) y el manto terrestre (magma), los

³ El grado en la que los errores de medición son disminuidos, es decir, que entre menos errores de medición existe una mayor validez (Villasís, Márquez, Zurita, Miranda y Escamilla, 2018).

⁴ La consistencia de las calificaciones en momentos distintos o en pruebas con ítems similares (Reidl, 2013).

cuales están integrados en una misma instrucción, aunque se trata de estructuras distintas una de la otra. Para este caso, es recomendable construir dos ítems: uno para preguntar sobre el ozono (las moléculas compuestas por tres átomos de oxígeno) y otro para el magma.

Una práctica apropiada para redactar ítems que integran aspectos, conceptos o contenidos distintos consiste en preguntar con base en las relaciones que guardan entre sí, tal como se muestra en el ejemplo a continuación:

Dos terceras parte del planeta Tierra están cubiertas de agua y, debido a su peso, la exploración de las fosas marinas más profundas es sumamente complicada. Las afirmaciones anteriores se refieren a las siguientes propiedades de las masas oceánicas:

- a) Masa y profundidad
- b) Volumen y densidad
- c) Composición química y masa

Como se observa, aunque se retoman dos elementos distintos en el ítem (peso del agua y la exploración submarina), los mismos tienen un aspecto en común: masas oceánicas, de modo que las personas pueden sojuzgar ambos elementos con base en un criterio común, siendo que para este caso la respuesta correcta es la opción “b”.

Error 2: Instrucciones dispersas o desconcentradas

Otro error se relaciona a incorporar instrucciones en las respuestas en lugar del encabezado del ítem. A continuación, se presenta un ejemplo de un ítem inadecuado en este sentido, según lo plantean Muñiz, Fidalgo, García, Martínez y Moreno (2005, pp.26-27).

En la investigación de referencia

- a) se controlan la variable A y la variable B
- b) se controla la variable A pero no la variable B
- c) se controlan la variable A y la variable B pero no la C

Como puede observarse, cada opción de respuesta tiene información que, en la buena práctica, debe estar en el encabezado. Así las cosas, un planteamiento adecuado del mismo ítem es el siguiente:

En la investigación de referencia, ¿cuáles variables se controlan?

- a) A y B
- b) A pero no B
- c) A y B pero no C

Como puede observarse en este caso, el encabezado del ítem contiene la información necesaria para contextualizar la pregunta, mientras que las opciones de respuesta son concretas y guardan relación a las instrucciones.

Error 3: Distractores heterogéneos o incoherentes

Este error consiste en incluir distractores que, durante la primera lectura no son coherentes con lo que se pregunta, lo cual repercute en la validez y confiabilidad del ítem.

Con base en lo anterior, un ítem redactado de forma inadecuada es el siguiente:

Refiere al punto más alto, en metros sobre el nivel del mar, de todo el relieve costarricense

- a) Monte Everest
- b) Cerro Chirripó
- c) Monte Fuji

Por las características de los distractores⁵ (opciones “a” y “c”), es sumamente evidente cuál es la respuesta correcta (opción “b”), de modo que un altísimo porcentaje de personas acertarán el ítem. Esto se debe a que las opciones incorrectas son heterogéneas.

⁵ Opciones incorrectas de respuesta (Jurado, Flores, Delgado, Sommer, Martínez y Sánchez, 2013).

Para mejorar lo anterior deben sustituirse los distractores por alternativas creíbles, tal como se presenta a continuación.

Refiere al punto más alto, en metros sobre el nivel del mar, de todo el relieve costarricense

- a) Volcán Irazú
- b) Cerro Chirripó
- c) Cerro de la Muerte

Como puede observarse, los distractores presentados son homogéneos con respecto a la opción correcta y se contextualizan en un mismo ámbito (corresponden a puntos altos del relieve costarricense), de modo que las personas deberán prestar más atención para seleccionar la alternativa correcta.

Error 4: Utilizar términos o expresiones que pueden generar sesgos involuntarios en las respuestas

El lenguaje influye en la percepción e interpretación que las personas tienen de los hechos que suceden a su alrededor. Muñiz, Fidalgo, García, Martínez y Moreno (2005) explican que algunos términos o expresiones tienden a asociarse con ciertas condiciones o significados, al sugerir que “términos o expresiones como ‘a veces’, ‘puede que’, ‘usualmente’ o ‘generalmente’ suelen ser percibidos por lo sujetos como asociados a contenidos ciertos, mientras que otros como ‘siempre’, ‘nunca’, ‘todos’ o ‘únicamente’ se asocian a contenidos falsos” (pp.27-28).

Asimismo, en esta tipología de error pueden incluirse juicios de valor que generalmente se relacionan al uso de términos tales como bueno, malo, mejor o peor, entre otros, así como proposiciones que tiendan a beneficiar o perjudicar a ciertas personas sin un propósito que le justifique, como por ejemplo afirmar que quienes aprenden a leer y a escribir antes de los cinco años tendrán dificultades para las matemáticas durante la adultez.

En este sentido, un ítem inadecuado puede presentarse de la siguiente forma:

¿Cuál recomendación debe tomarse en cuenta ante condiciones de tormenta eléctrica, con tal de resguardar la integridad física de las personas?

- a) Contar los segundos entre la luz y el sonido de los rayos
- b) Nunca nadar en piscinas cuando se escuchan truenos
- c) Leer un libro al aire libre mientras dejan de escucharse los truenos

Como puede observarse, la opción “b” resalta entre las demás por la palabra “nunca”, la cual podría generar dos situaciones: que la persona considere que dicha palabra le sugiere que es la respuesta correcta, lo cual haría muy transparente al ítem o, caso contrario, podría dudar y seleccionar alguna otra por duda.

Excepción con los ítems de escalamiento de Likert

Para los ítems de escalamiento de Likert, la inclusión de este tipo de palabras tiene consideraciones distintas, ya que pueden utilizarse para direccionarlo en uno u otro sentido (ver página 24 de este documento), tal como se explica a continuación.

Supóngase que se tiene la siguiente escala de Likert:

Totalmente de acuerdo	Algo de acuerdo	Indeciso	Algo en desacuerdo	Totalmente en desacuerdo
-----------------------	-----------------	----------	--------------------	--------------------------

Al momento de redactar los ítems para dicha escala, el uso de algunas palabras puede variar la intención o direccionalidad de los mismos, haciéndolos positivos o negativos, según se puede observar en los siguientes casos.

Caso No 1:

Considérese el siguiente ítem:

- Acaricio a los animales de pelaje corto que se acercan a mí.

De acuerdo a la afirmación anterior, las personas podrían contestar que están de acuerdo o no con la misma, dependiendo de sus gustos. Empero, el ítem podría cambiar su significado si se incorpora alguna palabra que varíe el contexto del mismo, tal como lo plantea el siguiente ejemplo:

- Acaricio únicamente a los animales de pelaje corto que se acercan a mí.

Como puede observarse, la inclusión de la palabra “únicamente” cambió el contexto del ítem sin invalidarlo y, por tanto, las respuestas de las personas podrían variar con base en este ajuste.

Caso No 2:

El ítem afirma lo siguiente:

- Me reconozco por respetar a las personas adultas mayores que conozco.

En el caso de este ítem, es muy probable que varias personas, quizá todas, contesten favorablemente ya que no se cuenta con un distractor y está planteado en forma positiva o favorable. Para proponer una redacción negativa del mismo que mejore la demanda y calidad de análisis que debe invertir una persona para contestarlo, se propone la siguiente redacción:

- Me reconozco por respetar solo a las personas adultas mayores que conozco.

Como puede verse, el sentido del ítem cambió en su significado, de modo que las personas deberán analizar con mayor cuidado su respuesta, ya que responder favorablemente a esta afirmación conllevaría a indicar que se respeta solo a las personas adultas mayores que la persona conoce, dejando por fuera a las demás, de modo que el ítem se ha replanteado negativo con respecto al presentado de primero y, por ello, requiere una lectura más cuidadosa para contestar.

En resumen, para el caso de los ítems con escalamientos de Likert, el empleo de este tipo de palabras puede ser beneficioso, siempre que sean utilizadas cuidadosamente y con la claridad adecuada.

Error 5: Utilizar opciones de respuesta con longitudes diferentes

Las opciones de respuesta deben tener una extensión similar. El mostrar opciones de respuesta con longitudes desiguales es un error que dificulta la comprensión adecuada del ítem en su totalidad, tal como se observa en la opción “c” del siguiente ejemplo:

Una de las causas relacionadas al aumento en los niveles de contaminación atmosférica es la siguiente:

- a) Actividad volcánica concentrada en regiones específicas
- b) Aumento de la actividad solar durante la última década
- c) Actividades antropogénicas que han demandado un consumo de combustibles fósiles que ha ido en significativo aumento durante las últimas décadas, causando una liberación excesiva de Dióxido de Carbono (CO₂) a la atmósfera

Error 6: Usos gramaticales incorrectos

El idioma y su uso son fenómenos dinámicos. Por ello, se recomienda redactar los ítems para que la mayoría de las personas los puedan comprender, tomando en cuenta el significado que ciertas palabras o términos tienen para algunos grupos etarios, niveles educativos o estratos generacionales, así como el uso adecuado del lenguaje inclusivo y la verificación de que exista una adecuada accesibilidad universal.

Algunas veces se cometen errores que pueden pasar por desapercibidos, pero que en el contexto de una prueba o test pueden generar brechas. En el siguiente ejemplo, tomado de Muñiz, Fidalgo, García, Martínez y Moreno (2005, p.28), existe una incongruencia con respecto al enunciado del ítem y la opción correcta:

El síndrome de Waterhouse-Friderichen está relacionado a la insuficiencia de cuál glándula:

- a) Las suprarrenales
- b) La pituitaria
- c) El páncreas

En este caso, las personas podrían sentirse confundidas por porque la redacción del encabezado del ítem y de las opciones “a” y “b” está en femenino, mientras que la opción “c”, la cual es la correcta, está escrita en masculino. Asimismo, existen otras inconsistencias tales como el uso del singular en el encabezado y en las opciones “b” y “c” de respuesta, pero la opción “a” está escrita en singular.

De esta forma, debe cuidarse la homogeneidad al momento de redactar cada ítem en todas las partes que le constituyen.

Error 7: Opciones de respuesta aleatorias o desordenadas

La simetría es una condición de constante búsqueda por parte del cerebro humano. Así las cosas, otro error consiste en mostrar las opciones de respuesta de forma desordenada cuando no existe ninguna razón que lo justifique.

Es recomendable que las opciones de respuesta se presenten en orden lógico, tanto la alternativa correcta como los distractores, tal como se presenta seguidamente.

El valor de pi (π) es

- a) 4,12
- b) 3,14
- c) 10,29

Es recomendable presentar las opciones de respuesta de forma ascendente (de menor a mayor) o descendente (de mayor a menor), tal como se muestra a continuación:

El valor de pi (π) es

- a) 3,14
- b) 4,12
- c) 10,29

Error 8: utilizar términos de connotación negativa

Es posible la sinceridad en la respuesta de los ítems sea un aspecto que se requiera medir en las pruebas. A veces esto se logra usando afirmaciones negativas que, como se ha visto, se trata de ítems con afirmaciones desfavorables que deben evaluarse de forma inversa.

Para el caso de los ítems con escalamientos de Likert se recomienda utilizar ítems positivos y negativos de forma balanceada, idealmente por mitades (Ej. 10 ítems positivos y 10 negativos por cada dimensión o constructo).

Conclusiones

La construcción de ítems constituye la materia prima de las pruebas que evalúan a las personas de forma objetiva. Por esta razón deben considerarse las recomendaciones que hace la literatura para llevar a cabo dicha tarea de forma adecuada.

Este informe presentó los distintos formatos de ítems a partir de sus dos tipos: ejecución máxima y ejecución típica, considerándolos con base en la evaluación de respuestas correctas e incorrectas y la evaluación por puntajes respectivamente. Asimismo, se ofrecieron ejemplos para cada caso, y se explicaron algunos de los errores más comunes que se cometen al redactarlos.

Si bien en este documento se presentaron insumos básicos que deben conocerse para la construcción de ítems, es importante aclarar que este tipo de procesos demanda de mucha creatividad, curiosidad e innovación, así como del desarrollo y puesta en práctica de habilidades lingüísticas, que permitan transformar ideas complejas en palabras simples, en los que se incluyan distintos elementos tales como imágenes, estudios de caso y otros similares como objetos de evaluación.

Todo lo anterior se conjuga con la mística que destaca a la mayoría de las personas que trabajan como funcionarias públicas, siendo que este tipo de documentos son la prueba de que existe un deseo permanente de hacer las cosas bien y cada vez mejor, dejando para ello legados que enriquezcan el orgullo que representa el privilegio de servir a la ciudadanía mediante las labores cotidianas.

Bibliografía

Enríquez, F. y Domínguez, A. (2010). Influencia de la Deseabilidad Social (DS) en Reportes de Capacitación. *Psicología Iberoamericana*, 18 (1), 69-79.

Haladyna, T., & Rodríguez, M. (2013). *Developing and validating test items*. New York: Routledge.

Hernández, R., Fernández, C. y Baptista, P. (2014). *Metodología de la investigación*. (6ª. ed.). Ciudad de México: McGraw-Hill.

Jurado, A., Flores, F., Delgado, L., Sommer, H., Martínez, A. y Sánchez, M. (2013). Distractores en preguntas de opción múltiple para estudiantes de medicina: ¿cuál es su comportamiento en un examen sumativo de altas consecuencias? *Investigación en Educación Médica*, 2(8), 202-210. ISSN 2007-5057.

Morales, P., Urosa, B. y Blanco, A. (2003). *Construcciones de escalas de actitudes tipo Likert*. Madrid: Editorial La Muralla, S.A.

Muñiz, J., Fidalgo, Á., García, E., Martínez, R. y Moreno, R. (2005). *Análisis de los ítems*. Madrid: La Muralla, S.A.

Reidl, L. (2013). Confiabilidad en la medición. *Investigación en Educación Médica*, 2(6), 107-111. ISSN: 2007-865X.

Rodríguez, M. (2003). Construct equivalence of multiple-choice and constructed-response items: A random effects synthesis of correlations. *Journal of Educational Measurement*, 40 (2), 163-184.

Villasís, M., Márquez, H., Zurita, J., Miranda, G. y Escamilla, A. (2018). El protocolo de investigación VII. Validez y confiabilidad de las mediciones. *Revista Alergia México*, 65(4), 414-421.