

Serie: Evaluación de aprendizajes

CÓMO DISEÑAR EVALUACIONES DE EXCELENCIA

Argumentos para usar preguntas de selección
múltiple para una evaluación precisa

Publicado por Aptus en marzo de 2024

Traducido por:
Aptus.org

Documento original de

CÓMO DISEÑAR EVALUACIONES DE EXCELENCIA

Argumentos para usar preguntas
de selección múltiple para una
evaluación precisa



Evidence Based
Education

¿POR QUÉ ES IMPORTANTE TENER EVALUACIONES DE EXCELENCIA?

Una evaluación efectiva es un proceso diseñado para generar inferencias o información práctica. Su objetivo es ayudar a los profesores a lograr objetivos claramente definidos que no podrían alcanzarse sin ella.

Aplicar una prueba o examen a menudo significa más que sólo evaluar conocimientos; las pruebas también pueden brindar oportunidades de aprendizaje (Roediger y Karpicke, 2006). Por ejemplo, una evaluación bien planificada le podría dar a usted la oportunidad de brindar retroalimentación de alta calidad sin aumentar su carga de trabajo.

Una forma de garantizar que la evaluación sea efectiva, es planificarla para que sea un puente sólido entre la enseñanza y el aprendizaje (William, 2013). Cuando usted dispone de una evaluación de excelencia, el proceso de determinar dónde están los estudiantes, dónde deben estar y cómo llevarlos allí se vuelve mucho más fácil!

¿POR QUÉ ES IMPORTANTE TENER EVALUACIONES DE EXCELENCIA?

Por supuesto, sabemos que hay muchas formas de evaluar, pero en esta guía nos enfocamos específicamente en el uso de preguntas de selección múltiple (PSM).

Por lo general, los profesores utilizan las PSM para evaluar conocimiento de datos y lo que se suele llamar habilidades de de orden inferior, que permiten evaluar si los niños recuerdan y entienden esos datos o hechos (Tarrant et al., 2006; Zimmaro, 2010). No obstante, cuando las PSM tienen un propósito y están bien planificados, pueden generar un aprendizaje de nivel superior y crear una dificultad deseable, lo que aumenta positivamente el valor de la evaluación (Coe, 2013).

Al igual que con cualquier herramienta educativa, se debe evaluar si las PSM se adaptan al propósito particular establecido antes de decidir usarlas.

BENEFICIOS Y LIMITACIONES DE LAS PSM

BENEFICIOS

- Producen información más confiable que otros tipos de pruebas, como las de preguntas abiertas, ya que se pueden calificar de forma objetiva (Burton et al., 1991).
- El estudiante puede responder más preguntas de selección múltiple que preguntas abiertas, por lo que cubre una base temática más amplia y representativa (Zimmaro, 2010; Burton et al., 1991).
- ¡Son rápidas de corregir y rápidas de completar! (Christodoulou, 2016).
- Una PSM bien elaborada permitirá que el profesor identifique errores conceptuales fácilmente y en la misma cantidad de tiempo (Christodoulou, 2016).

LIMITACIONES

- Pueden ser susceptibles a que el estudiante adivine la respuesta (Burton et al., 1991), aunque es poco probable que un estudiante destaque en una prueba utilizando esta técnica (Christadoulou, 2016).
- No pueden medir habilidades de creación como la capacidad de escritura, la creatividad o la articulación del pensamiento (Zimmaro, 2010).
- Son lentas de elaborar y una buena construcción de una PSM toma tiempo (Zimmaro, 2010).
- Los resultados de las PSM están influenciados por tendencias a tomar riesgos. Investigaciones muestran que los niños suelen tratar de adivinar más las respuestas que las niñas, y esta toma de riesgos también se asocia con ciertas culturas y antecedentes socioeconómicos (Ben-Shakhar y Sinai, 1991).

¿CÓMO DISEÑAR UNA PRUEBA DE SELECCIÓN MÚLTIPLE?

Ahora que se hemos considerado la evidencia, debemos pensar en cómo diseñar una prueba de selección múltiple.

Antes de escribir las preguntas de una prueba de este tipo, debemos estar seguros de que no solo estamos evaluando una pequeña parte del contenido en función de los objetivos de aprendizaje clave de nuestro currículum.

Una forma de hacer esto es crear una tabla basada en constructos u objetivos de aprendizaje de nuestros planes de estudio. Este es un método sistemático que ayuda a identificar tanto los objetivos de aprendizaje que se deben cubrir como la complejidad de la comprensión que se quiere evaluar.

En la siguiente tabla, un docente utilizó una matriz para diseñar una prueba de selección múltiple que evalúa si los estudiantes pueden identificar e interpretar el uso de verbos en tiempo presente. Puede parecer complejo al principio, pero puedes seguir la guía paso a paso para comprender el proceso.

MATRIZ PARA UNA EVALUACIÓN DE FRANCÉS PARA TERCER GRADO

Objetivo de aprendizaje	Identificar/ Nombrar	Interpretar/ Describir/ Enumerar	Ponderación	N.º esperado de preguntas (por objetivo de aprendizaje)
Verbos regulares con "er" en tiempo presente	2	2	40 %	4
Verbos irregulares con "er" en tiempo presente	1	1	20 %	2
Verbos con "ir" en tiempo presente	1	1	20 %	2
Verbos con "re" en tiempo presente	1	1	20 %	2
Ponderación	50 %	50 %	100 %	
Nº esperado de preguntas (por habilidad)	5	5		

La ponderación y el número esperado de preguntas dependen totalmente de nosotros y deben basarse en las necesidades de los estudiantes. Este modelo refleja una evaluación que tiene el propósito de evaluar la comprensión básica de los conceptos gramaticales clave del estudiante, en lugar de sus habilidades de creación.

Hay que recordar que esta matriz se puede usar para planificar y ponderar cualquier tipo de evaluación —desde aquellas informales hasta las pruebas escritas del final de la unidad, pasando por las preguntas bisagra que podemos hacer verbalmente durante las clases—. Dicho esto, aquí le mostramos cómo hacer la suya.

PASOS PARA CREAR UNA MATRIZ

Paso 1: decidir cuántas preguntas se necesitan para lograr el propósito de la prueba. Se debe determinar el número de preguntas para que se ajuste a la cantidad óptima de tiempo que está dispuesto a dedicar a la evaluación, ya sean segundos u horas. Una prueba de diez preguntas otorga una representación justa de los constructos a evaluar y puede darle más ponderación a un constructo que es más fundamental que otros.

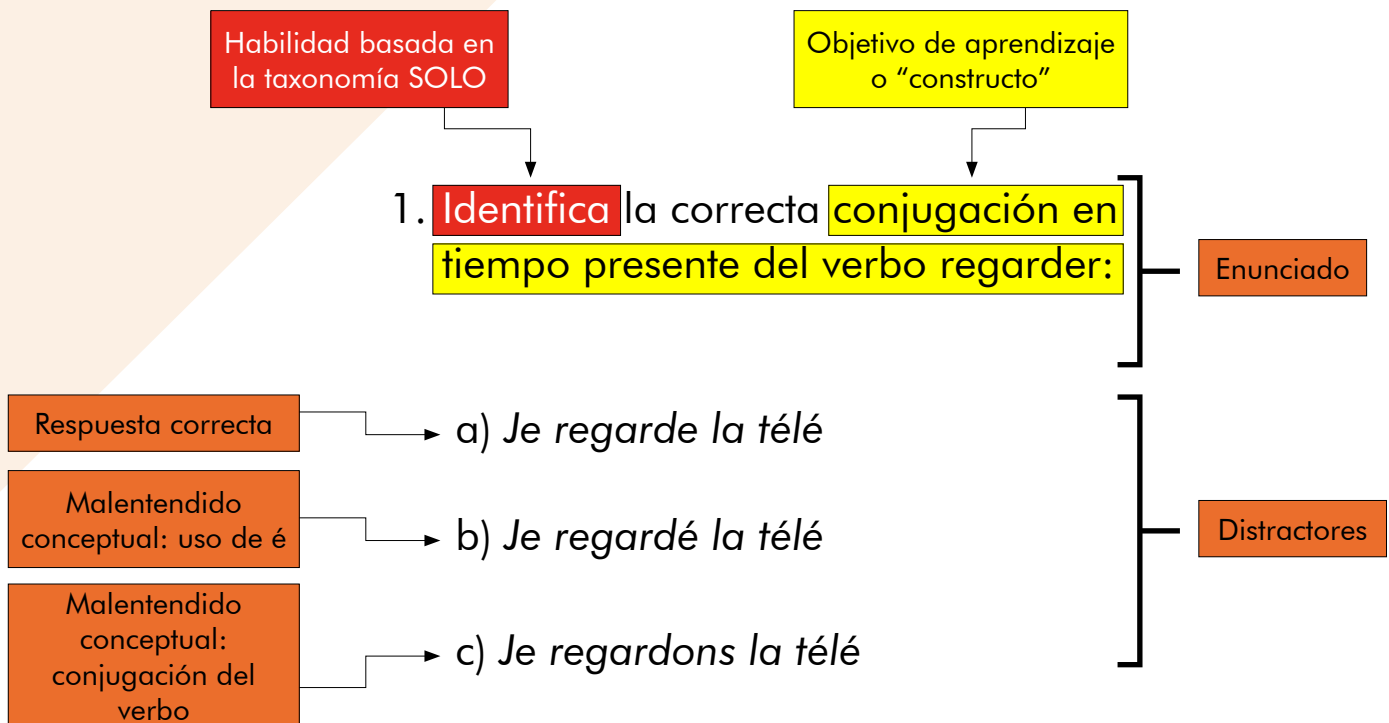
Paso 2: decidir qué tipo de habilidades son útiles para acceder al contenido de la evaluación y escribirlas en el eje X. Estas habilidades deben reflejar la profundidad de la comprensión siguiendo una taxonomía (por ejemplo, la de Bloom o SOLO —*Structure of Observed Learning Outcome*—). En este ejemplo, el propósito es evaluar la comprensión básica que tiene el estudiante de conceptos gramaticales clave en tiempo presente, en lugar de sus habilidades de creación. Por lo tanto, los constructos elegidos son identificar, interpretar y analizar.

Paso 3: agregar lo que queremos evaluar. Los ítems a lo largo del eje Y son nuestros objetivos de aprendizaje o constructos. Estos deben tomarse del plan de estudios derivado del currículum. La lista solo debe incluir los ítems que se quieren evaluar que han sido seleccionados porque se podrá **hacer algo** con la información generada; cualquier otra cosa es irrelevante para el propósito de esta prueba. El profesor del ejemplo quería enfocarse en la comprensión de los verbos con “er”, ya que ese había sido el foco en clases anteriores, al mismo tiempo que revisaba otros verbos regulares.

Paso 4: ingresar un porcentaje que indique cuánta ponderación queremos que tenga cada objetivo de aprendizaje y habilidad. Esto depende del enfoque que necesitemos para nuestro trabajo. ¡Lo principal aquí es asegurarse de que las ponderaciones sumen un total de 100 %!

[adaptado de Rodriguez y Albano, 2017].

¿CÓMO DISEÑAR PREGUNTAS DE OPCIÓN MÚLTIPLE DE ALTA CALIDAD?



- Las PSM constan de dos partes: el enunciado y las alternativas.
- El enunciado debe ser comprensible (Brame, 2013) y debe incluir una pregunta que estimule las respuestas y se centre en resultados de aprendizaje específicos.
- Las alternativas incluyen la respuesta correcta o clave y las incorrectas.
- Las alternativas deben tener un contenido similar (Boland et al., 2010; Burton et al., 1991).
- La respuesta correcta no debe ser más larga ni más detallada, ya que en ocasiones puede interpretarse como una pista (Tarrant y Ware, 2008).
- El vocabulario debe ser fácil de comprender (Burton et al., 1991; Moreno et al., 2006).
- Las preguntas deben ser independientes entre sí, es decir, la siguiente pregunta no debe requerir responder correctamente la anterior (Brame, 2013).
- Se debe evitar formular las preguntas con frases negativas, por ejemplo, "cuál de estos **no** es un..." (Burton et al., 1991).
- Evitar usar palabras como "nunca", "siempre", "generalmente" (Boland et al., 2010).

LISTA DE VERIFICACIÓN DE PREGUNTAS DE SELECCIÓN MÚLTIPLE

Lista de verificación	SÍ	NO
Antes de escribir los ítems, los objetivos de aprendizaje se ponderan en función de una matriz de evaluación.		
Cada pregunta evalúa solo un objetivo de aprendizaje bien identificado.		
Cada pregunta es independiente una de la otra.		
La evaluación no utiliza vocabulario complicado que es innecesario.		
No hay preguntas basadas en opiniones.		
Las preguntas fáciles se ubican al comienzo de la prueba.		
Se usan letras delante de las alternativas, no números.		
Todos los enunciados están redactados de la manera más corta posible.		
Todos los enunciados solo incluyen contenido relevante.		
Los enunciados no están redactados de forma negativa.		
Los enunciados no dan pistas ni excluyen algunas de las alternativas.		
Las alternativas son mutuamente excluyentes y no se superponen.		
Todas las alternativas se presentan en orden lógico.		
Todas las alternativas están redactadas de manera similar.		
Solo hay una respuesta correcta en cada pregunta.		
La opción correcta no ofrece más detalles, ni es más larga.		
Todos los distractores son posibles.		
Los distractores se basan en malentendidos conceptuales de los estudiantes.		

PASOS SIGUIENTES

Al momento de analizar los datos de las respuestas de los estudiantes, el uso efectivo de las alternativas ayudará a identificar errores conceptuales en su aprendizaje, lo cual es tan útil como saber qué respondieron correctamente. Si menos del 5 % de los estudiantes eligió un distractor, podría ser recomendable reconsiderar si ese distractor se debe eliminar o modificar (Tarrant et al., 2009; Towns, 2014).

Una vez que se identifican los errores conceptuales específicos, podemos dirigir a los estudiantes al ejercicio complementario pertinente, ya sea repasando aprendizajes previos o pasando a una tarea de “estiramiento y desafío” para aquellos que sobresalieron. Usar las evaluaciones de esta manera para la evaluación formativa las vuelve significativas para los estudiantes y las convierte en un ejercicio de recopilación de datos mucho más rápido y confiable para usted.

El uso de las PSM como preguntas de recuperación, preguntas bisagra, pruebas y en la evaluación para el aprendizaje se puede explorar más en la Evidence Based Education’s Assessment Academy (Academia de Evaluación de la organización Evidence Based Education). La Assessment Academy cierra la brecha entre la academia y la sala de clases para ayudar a los profesores a evaluar mejor. Más información en evidencebased.education

GLOSARIO DE TÉRMINOS

Alternativas: opciones proporcionadas por las preguntas de opción múltiple, que incluyen tanto las respuestas correctas como las incorrectas.

Matriz (matriz de evaluación): plan o mapa escrito de la especificación de evaluación que relaciona los temas principales con las habilidades cognitivas y la cantidad de preguntas evaluadas.

Dominio: área.

Distractores: alternativas incorrectas proporcionadas como opciones en una pregunta de opción múltiple.

Ítem: pregunta en un examen.

Clave: opción correcta en una pregunta de opción múltiple.

Equilibrio de claves: técnicas sistemáticas utilizadas para mejorar los patrones en que aparece una respuesta correcta en una prueba.

Error conceptual: visión incorrecta sobre un tema debido a una comprensión defectuosa.

Enunciado: parte de la pregunta de opción múltiple que plantea la pregunta y exige una respuesta.

Sabiduría para contestar pruebas (*testwiseness*): comportamiento que permite que los examinados aumenten su puntaje en la prueba, aunque no sepan la respuesta a un ítem de opción múltiple. Esto puede ocurrir debido a adivinar o deducir una respuesta correcta en función de pistas en la pregunta (Downing, 2002).

REFERENCIAS

- Ben-Shakhar, G., & Sinai, Y. (1991) Gender differences in multiple-choice tests: the role of differential guessing tendencies. *Journal of Educational Measurement*, 28(1), 25-35.
- Boland, R., Lester, N., & Williams, E. (2010). Writing multiple-choice questions. *Academic Psychiatry*, 34(4), 310-316.
- Brame, C. (2013). Writing good multiple-choice test questions. Recuperado de <https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test.questions>.
- Burton, S. J., Sudweeks, R. R., Merrill, P. G., & Wood, B. (1991). *How to Prepare Better MultipleChoice Test Items: Guidelines for University Faculty*. Provo, UT: Brigham Young University Testing Services and The Department of Instructional Science.
- Coe, R. (2013). [online] Cem.org. Disponible en: <http://www.cem.org/attachments/publications/ImprovingEducation2013.pdf> [Recuperado el 13 de marzo de 2019]. Moreno, R., Martínez, R. J., & Muñoz, J. (2006). New guidelines for developing multiple-choice items.
- Christodoulou, D. (2016). *Making good progress? The future of Assessment for Learning*. Oxford, England: Oxford University Press.
- Methodology: European journal of Research Methods for the Behavioral and Social Sciences*, 2(2), 65-72. Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2), 3-13.
- Roediger, H. and Karpicke, J. (2006). *Test-Enhanced Learning*. [online] St Louis: Washington University, pp.249-255. Disponible en: http://learninglab.psych.purdue.edu/downloads/2006_Roediger_Karpicke_PsychSci.pdf [Recuperado el 13 de marzo de 2019].
- Tarrant, M., Knierim, A., Hayes, S. K., & Ware, J. (2006). The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. *Nurse Education in Practice*, 6(6), 354-363.
- Towns, M. H. (2014). Guide to Developing High-Quality, Reliable, and Valid Multiple-Choice Assessment. *Journal of Chemical Education*, 91(9), 1426-1431.
- William, D. (2006). *Assessment: The Bridge between Teaching and Learning*. Voices from the Middle, 21(2), pp.15-20.
- Zimmaro, D. M. (2010). *Writing Good Multiple-Choice Exams*. Recuperado de: <https://ctl.utexas.edu/assets/Evaluation--Assesment/Writing-Good-Multiple-Choice-Exams-04-28-10.pdf>