



## Los Efectos Posicionales en Ítems de Opción Múltiple: el Orden de los Distractores Afecta a la Dificultad

Irene Serrano-Blaya y Carmen García

Departamento de Psicología Social y Metodología, Facultad de Psicología, Universidad Autónoma de Madrid, España

### INFORMACIÓN DEL ARTÍCULO

*Historia del artículo:*  
Recibido 28 July 2023  
Aceptado 17 April 2024

*Palabras clave:*  
Ítems de opción múltiple  
Posición del distractor  
Posición de la opción correcta  
Orden de las opciones

### RESUMEN

La creciente aplicación de evaluaciones con pruebas de opción múltiple acentúa la importancia de medir rigurosamente. Está demostrado el efecto posicional de las opciones de respuesta del ítem sobre sus propiedades psicométricas (dificultad y discriminación), aunque la evidencia disponible es limitada y contradictoria en algunos puntos. La posición de la opción correcta ha sido una primera línea de estudio, mientras que trabajos recientes han analizado la posición de los distractores u opciones incorrectas, según su plausibilidad o atractivo. En este trabajo se evalúa el efecto posicional de la opción correcta y el distractor más y menos elegido, sobre la dificultad y discriminación de ítems de tres, cuatro y cinco opciones. Se han hallado resultados relacionados con el "efecto anclaje" en la creación y distribución de las opciones: los distractores atractivos colocados al final dificultaban el ítem, mientras que los menos elegidos seguían la tendencia contraria. No hubo efecto en la discriminación.

### Positional effects on multiple choice items: Distractor order affects item difficulty

### ABSTRACT

The increasing popularity of multiple-choice assessment implies the need for accurate evaluation. The available literature shows a positional effect of response options on items' psychometric properties (item difficulty and discrimination), albeit evidence is scarce and often conflicting. Historically, correct answer position was firstly explored, followed by distractor position, which is a novel line of research focusing on distractor plausibility or attractiveness. In this study, response-option positional effects on item difficulty and discrimination were assessed, using three, four and five-response-option items, regarding the position of the correct answer, and the most and least chosen distractors. Distractor position affected item difficulty, coinciding the "anchoring effect" in response option distribution: when the most attractive distractor was placed at the last positions, items became more difficult, and the opposite effect was found for the least chosen one. Distractor position had no effect on item discrimination.

Los exámenes de opción múltiple (OM), también conocidos como "tipo test", se popularizaron desde principios del siglo XX para mejorar la evaluación, por razones como una mejor relación coste-eficacia y permitiendo evaluar más contenidos de una vez o incluso permitiendo diagnosticar fallos en el aprendizaje de los examinados (Gierl et al., 2017). Su amplio uso en la actualidad precisa de estudios que garanticen una correcta elaboración, aplicación y corrección para así reducir las desigualdades en la evaluación, y hacer estas lo más fidedignas que sea posible.

Los exámenes de opción múltiple (OM) se componen de varios ítems, cada uno de los cuales tiene un enunciado y varias opciones de respuesta (OOR), entre ellas la opción correcta (OC), y distractores

u opciones incorrectas, que teóricamente deberían atraer a los estudiantes con menos conocimientos para distinguir entre los distintos niveles de habilidad.

Para asegurar una evaluación justa con este tipo de instrumentos es necesario prestar atención tanto al proceso de construcción de los ítems como al análisis de las propiedades psicométricas del test, los ítems y las distintas OOR. El grupo de investigación de Haladyna y colaboradores (Haladyna y Rodríguez, 2013) propone una serie de directrices para la elaboración de los ítems de OM, a fin de minimizar la introducción de varianza no relacionada con el constructo que se pretende medir, evitando así posibles variaciones de los resultados no relacionadas con el nivel de habilidad de los estudiantes del tema

Para citar este artículo: Serrano-Blaya, I. y García, C. (2024). Los efectos posicionales en ítems de opción múltiple: el orden de los distractores afecta a la dificultad. *Psicología Educativa*, 30(2), 111-118. <https://doi.org/10.5093/psed2024a10>

Correspondencia: [ireneserranoblaya@gmail.com](mailto:ireneserranoblaya@gmail.com) (I. Serrano-Blaya).

ISSN: 1133-0740/© 2024 Colegio Oficial de la Psicología de Madrid. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

evaluado, o conocimiento del mismo. Estas directrices evitan fenómenos como el *testwiseness* (Haladyna y Rodríguez, 2013), que es la facilidad que tienen algunos examinados para averiguar la OC con estrategias que van más allá de su nivel de conocimiento, que se basan en la capacidad de identificar determinadas claves y patrones en los ítems de OM. Además, cuidar la construcción de ítems con este tipo de directrices contribuye a que sus propiedades psicométricas sean adecuadas. Cabe decir, sin embargo, que las directrices no siempre se aplican de forma generalizada (Lions et al., 2022).

Las propiedades psicométricas, por otra parte, dan información sobre si en la evaluación se diferencia eficazmente a los estudiantes con un grado elevado o bajo de conocimiento o habilidad. Con esa finalidad se revisan en el test completo la fiabilidad o precisión en la medida y su validez. Además se puede evaluar la calidad de los ítems con indicadores como la dificultad –medida con el Índice de Dificultad, o proporción de aciertos ( $p_j$ )– y la discriminación –medida con el Índice de Discriminación ( $D_j$ ) o con la correlación ítem-resto del test–, que nos informa de la capacidad del ítem para diferenciar entre examinados con alto y bajo nivel de conocimiento. La plausibilidad es otro indicador de buen funcionamiento de los ítems, refiriéndose a la capacidad de un distractor de ser elegido por los examinados, es decir, señala en qué medida esta opción resulta una alternativa creíble para aquellos que no saben cuál es la respuesta correcta. Normalmente, para que un distractor se considere plausible, se establece el punto de corte del 5% de elecciones recibidas de un ítem. No obstante, en ítems muy fáciles, en los que casi todos los examinados seleccionan la OC, es difícil que cualquier distractor llegue a ese punto de corte, incluso cuando podría ser una opción creíble para los que no saben la respuesta correcta. Este es el motivo por el que Raymond et al. (2018) proponen el nuevo criterio de plausibilidad,  $p_{\text{net}}$ , que varía en la medida en que varía la dificultad del ítem, evitando así este problema. Por otro lado, cabe decir que, según Tarrant et al. (2009), los ítems no suelen tener más de dos distractores plausibles, razón por la cual algunos autores (Haladyna et al., 2019; Rodríguez, 2005) recomiendan que se usen preferentemente los ítems de tres OOR. Este tipo de ítems son más fáciles de elaborar, tienen unas propiedades psicométricas al menos tan adecuadas como los ítems con más opciones, además de ser más eficientes en la evaluación, porque permiten evaluar más contenidos en el mismo tiempo y mejorar la representatividad del constructo (Holzknecht et al., 2020; Rodríguez, 2005).

De acuerdo con la literatura, otra variable que podría tener un efecto en los indicadores psicométricos y la calidad de los ítems (dificultad y discriminación) es la posición de las OOR, introduciendo fuentes de error en la medida y afectando a las puntuaciones de los examinados, como resultado de las diferencias en el procesamiento de la información (Shin et al., 2019), de la tendencia a colocar o a seleccionar la OC en ciertas posiciones (Attali y Bar-Hillel, 2003), o a una habilidad especial para resolver exámenes de OM, como el *testwiseness*. Según Carnegie (2017), la posición de las OOR puede dar pistas a los examinados, aumentando su probabilidad de acierto.

## Los Efectos de la Posición de las Opciones de Respuesta en la Dificultad y Discriminación del Ítem

### Los Efectos de la Posición de la Opción Correcta

La mayoría de la investigación sobre efectos posicionales se centra en la OC y en cómo afecta su localización a la dificultad y discriminación. En ítems sin significado, en los que se elige al azar, como en una quiniela, se ha descrito una tendencia a seleccionar las opciones centrales (Wilbur, 1966), lo que no está tan claro en evaluaciones reales y en ítems con significado. Cronbach (1950) fue pionero en el estudio de los efectos de la posición en los exámenes de OM, concluyendo que el efecto de estos sesgos sería despreciable. Posteriormente, auto-

res como Cizek (1994) y Marcus (1963) encuentran datos parecidos y concluyen que los sesgos de posición son un fenómeno variable y que no tendrían consecuencias psicométricas serias.

Sin embargo, algunos autores sí han encontrado algunos sesgos con respecto a la posición de la OC. En varios trabajos se observa un “sesgo de primacía”, término acuñado por Tellinghuisen y Sulikowski (2008), que supone la preferencia por una primera opción cuando hay dos posibilidades igualmente atractivas. Bresnock et al. (1989) proponen que cuantos más distractores lee el examinado antes de encontrar la OC mayor es la confusión. Del mismo modo, Schroeder et al. (2012) explican que aumentaría progresivamente la carga cognitiva según se leen más distractores antes de la OC, lo que podría explicar el fenómeno descrito. En un estudio reciente (Hagenmüller, 2020) se encuentra un efecto parecido al sesgo de primacía en evaluaciones de la inteligencia, con ítems más fáciles cuando la OC se colocaba en primeras posiciones y más difíciles cuando se colocaba en el medio o al final. Mediante la metodología de *eye tracking*, Holzknecht et al. (2020) observaron que los evaluados dedicaban más tiempo a leer la primera opción, y progresivamente menor tiempo a medida que se pasaba a las subsiguientes, lo que podría hacer que los examinados no leyeran el resto de opciones si encontraban la OC o un distractor convincente en las primeras posiciones, especialmente si la limitación del tiempo en la evaluación era importante.

Por otro lado, Attali y Bar-Hillel (2003) hallaron un sesgo central, por el cual tanto examinadores como examinados tenderían tanto a colocar como a seleccionar la OC en las opciones centrales y a evitar las de los extremos, fenómeno que denominaron “aversión a los extremos”. No obstante, este sesgo no se ha replicado en la literatura posterior. Por último, a favor de la existencia de sesgos posicionales, nos encontramos con la posibilidad de inducir experimentalmente una preferencia de posición en los examinados. Wevrick (1962), con el método de “reforzamiento parcial de una posición”, consiguió que los estudiantes prefiriesen la segunda posición de ítems de cuatro opciones, de manera que, cuando la OC se movía a otra posición, los ítems resultaban más difíciles. Por el contrario, Carnegie (2017) no encontró demostración de que una serie de ítems con la OC en la misma posición afectara a la puntuación de los estudiantes.

### Los Efectos de la Posición de los Distractores

Esta línea de investigación es más reciente que la anterior (Gierl et al., 2017) y se ha centrado en la posición de aquellos distractores más plausibles y atractivos, que son los que más efectos posicionales han demostrado, hablándose en este caso de los llamados distractores altamente plausibles (DAP) y del distractor más atractivo (DMA) de un ítem. Aún más escasos son los estudios sobre los efectos de posición de los distractores menos plausibles, con resultados aún preliminares.

En este caso, también se describe un efecto de primacía (Schroeder et al., 2012), según el cual los distractores se eligen más en las primeras posiciones y que los DAP colocados al principio generan mayor confusión y merman la proporción de acierto. Por su parte, Shin et al. (2019) observaron que los ítems eran más difíciles cuando el DMA precedía a la OC respecto al orden inverso, y que el efecto era mayor cuanto más cerca estuvieran las dos opciones. En este trabajo, los estudiantes con un elevado nivel de conocimiento resistieron mejor efectos posicionales que los estudiantes de bajo nivel de conocimiento. En ese mismo sentido, Kiat et al. (2018) encuentran que independientemente del nivel de habilidad del examinado, la plausibilidad del distractor tenía un efecto moderador en la relación entre la posición y la elección del ítem: los distractores plausibles colocados antes de la OC se seleccionaban más frecuentemente que en el orden inverso.

En el estudio de Lions et al. (2021) se encontró que los distractores más plausibles se colocaban con mayor frecuencia en posiciones

centrales y los menos creíbles en las últimas posiciones, concluyendo así que los mejores distractores suelen ir antes que los peores. Los distractores tienden a generarse en orden de plausibilidad porque son los más fácilmente accesibles, fenómeno que [Attali y Bar-Hillel \(2003\)](#) denominaron “efecto anclaje” (*anchoring effect*). [Lions et al \(2021\)](#) postulan que podría ser la posición de los distractores más plausibles, y no de la OC, lo que produzca el sesgo central de respuesta, de manera que la posición de los distractores pudiera ser un factor contaminante, que no se suele tener en cuenta en la investigación sobre este tema.

En relación con los distractores menos plausibles, [Kiat et al. \(2018\)](#) encuentran que la posición de estos no influye en la dificultad del ítem en su test. Por otro lado, [Lions et al. \(2021\)](#) postulan que al colocarse estas opciones al final pueda ocurrir que los estudiantes las descarten rápidamente, aumentando su probabilidad de acertar.

### Los Efectos de la Aleatorización y el Reparto Aproximadamente Uniforme de la Posición de las Opciones

Para evitar fuentes de error relacionadas con los efectos posicionales se pueden reorganizar las OOR, para lo cual contamos con tres métodos principales. El primero sería el reparto aproximadamente uniforme de la OC, en el que esta opción se distribuiría de manera equiprobable en las distintas posiciones. La práctica habitual es prestar atención exclusivamente a la posición de la OC, pero algunos autores recomiendan reorganizar todas las OOR, habida cuenta de los efectos posicionales de los distractores encontrados en la literatura ([Shin et al., 2019](#)). En segundo lugar, se podría aleatorizar la posición de las OOR, que difiere del anterior método en que se usa un método externo para reordenar las opciones, como por ejemplo un dado. Una tercera opción sería la aleatorización del orden de las preguntas del test. A su vez, la aleatorización puede aplicarse por igual para todos los examinados, por grupos o individualmente. Estas dos últimas opciones reducen la posibilidad de que los estudiantes se copien las respuestas, pero podría suponer cambios impredecibles e injustos en las puntuaciones como consecuencia de los efectos de posición ([Cizek, 1994](#)).

Como argumentos a favor de la reorganización de las OOR nos encontramos con la posibilidad de inducir sesgos posicionales si se refuerzan ciertas alternativas de respuesta ([Wevrick, 1962](#)), así como el “sesgo de anclaje” propuesto por [Attali y Bar-Hillel \(2003\)](#), según el cual los examinadores tienden a producir las alternativas de respuesta en orden decreciente de plausibilidad, de manera que de no reordenarse los distractores, los menos plausibles quedarían siempre al final, lo cual podría tener implicaciones sobre la dificultad y discriminación del ítem. Otros autores sugieren cautela a la hora de reordenar las opciones ante la posibilidad de inducir nuevos efectos posicionales, como por ejemplo el efecto primacía, de manera que algunas versiones sean más fáciles que otras por azar ([Bresnock et al., 1989](#)).

En conclusión, se recomienda utilizar algún tipo de redistribución del orden de las OOR. El reparto aproximadamente uniforme de la OC se considera una práctica adecuada, aunque sería preferible que se incluyesen además los distractores ([Gierl et al., 2017](#)). Otros autores, por el contrario, defienden el uso preferente de la aleatorización, por su mayor rapidez y para evitar nuevos sesgos propios de los desarrolladores y las estrategias de *testwiseness* por compensación de las posiciones de la OC a lo largo del test ([Lions et al., 2022](#)). No obstante, es cierto que cualquier reorganización posicional puede a su vez inducir nuevos sesgos y afectar a las propiedades psicométricas del ítem ([Hagemüller, 2020](#)), razón por la cual la reorganización aplicada de manera distinta para cada estudiante quedaría desaconsejada, y por lo que se recomienda analizar el funcionamiento de los ítems siempre que se modifique la posición de sus opciones.

En resumen, la literatura disponible muestra evidencias de efectos de posición en exámenes de OM, en algunos casos contradictoria y en otros escasa. Tal es el caso de las recientes líneas de investigación relativas a los efectos de la posición de los distractores altamente plausibles (DAP), así como aquellos con baja plausibilidad. Además, no se conocen estudios que comparen los efectos de posición en ítems de tres, cuatro y cinco OOR, información que podría ayudar a establecer recomendaciones útiles para los desarrolladores de exámenes. El presente estudio pretende dar respuesta a carencias en la investigación, examinando los efectos ampliamente analizados –de la posición de la OC, aquellos que aparecen en nuevas líneas de investigación –sobre la posición de distractores de alta y baja plausibilidad–, todo ello comparando ítems con distinto número de OOR. Los resultados de este estudio presentarían una contribución única, con consecuencias importantes para la evaluación con exámenes de OM.

### Objetivos

El objetivo de este estudio es examinar los efectos de la posición de la OC y de los distractores en los indicadores de dificultad y discriminación del ítem. Se analizarán los efectos posicionales de la opción correcta (OC), el distractor más atractivo (DMA) y el distractor menos elegido (DME) sobre estos indicadores, comparando ítems de OM de tres, cuatro y cinco OOR.

### Método

#### Base de Datos

La base de datos utilizada en este estudio está compuesta por 64 exámenes y 1,989 ítems de OM de tres, cuatro y cinco alternativas de respuesta, provenientes de distintos grados universitarios y de varias materias impartidas en la Universidad Autónoma de Madrid (UAM).

Más específicamente, la base de datos estuvo conformada por 24 exámenes y 754 ítems de tres OOR, 23 exámenes y 727 ítems de cuatro opciones, y 17 exámenes y 508 ítems de cinco opciones. Cada examen contaba con al menos 70 respuestas. La calidad de los ítems y sus propiedades psicométricas se evaluaron con diversos estadísticos, separando por tipo de ítem (tres, cuatro o cinco OOR), tal como se observa en la [Tabla 1](#). El coeficiente alfa de Cronbach del test completo –de cada examen por separado– alcanzó una media de .697 y una desviación típica de .127.

#### Análisis de Datos

Para obtener los datos relativos a la varianza, dificultad y discriminación del ítem, porcentaje de elecciones de cada opción, así como los estadísticos del test, se utilizó el programa TAP ([Brooks y Johanson, 2003](#)). El análisis de datos se llevó a cabo con el programa SPSS V.25 y se usaron las pruebas ji-cuadrado y ANOVA. La primera de ellas, para comprobar si la distribución de las distintas opciones (OC, DME, DMA), era homogénea. Por su parte, la prueba ANOVA se utilizó para observar el efecto de la posición de las distintas opciones (OC, DME, DMA) en la dificultad y discriminación del ítem.

En cuanto a la prueba ji-cuadrado, tras obtener un valor significativo al 99% de confianza, el criterio de decisión, en cuanto a qué frecuencias observadas diferían significativamente de las esperadas, fue comprobar que los límites superior e inferior del intervalo de confianza no encerrasen el valor de proporción esperada ([Pardo y San Martín, 2015](#)), que sería del 33, 25 y 20% para los ítems de tres, cuatro y cinco OOR, respectivamente.

Con respecto a la prueba ANOVA, como se realizaron múltiples comparaciones, se utilizó un intervalo de confianza del 99% para reducir la probabilidad de cometer un error Tipo I ([Pardo y San Martín, 2015](#)). También se utilizó dicho intervalo de confianza para las com-

**Tabla 1.** Estadísticos Descriptivos de los Indicadores Psicométricos de los Ítems Para Ítems de Tres, Cuatro y Cinco Opciones de Respuesta

Nº OOR	Indicador	Mínimo	Máximo	Media	DT
3	Índice de Dificultad ( $p_j$ ) <sup>1</sup>	.010	1.000	.642	.230
	Índice de Discriminación ( $D_j$ ) <sup>1</sup>	-.340	.890	.299	.120
	Correlación ítem-resto del test <sup>2</sup>	-.360	.620	.223	.137
	Varianza <sup>1</sup>	.000	.250	.177	.065
4	Índice de Dificultad ( $p_j$ ) <sup>3</sup>	.030	1.000	.644	.242
	Índice de Discriminación ( $D_j$ ) <sup>3</sup>	-.190	.870	.295	.180
	Correlación ítem-resto del test <sup>4</sup>	-.330	.590	.220	.146
	Varianza <sup>3</sup>	.000	.250	.170	.074
5	Índice de Dificultad ( $p_j$ ) <sup>5</sup>	.050	1.000	.713	.223
	Índice de Discriminación ( $D_j$ ) <sup>5</sup>	-.090	.700	.271	.156
	Correlación ítem-resto del test <sup>6</sup>	-.260	.660	.222	.139
	Varianza <sup>5</sup>	.000	.250	.155	.075

Nota. OOR = opciones de respuesta.

<sup>1</sup>N = 754; <sup>2</sup>N = 751; <sup>3</sup>N = 727; <sup>4</sup>N = 715; <sup>5</sup>N = 608; <sup>6</sup>N = 502.

paraciones de tendencia o la prueba Tukey de comparaciones *post hoc*.

Por su parte, se calcularon los índices  $p_{nfd}$  (Raymond et al., 2018) de plausibilidad de los distractores y omega cuadrado de tamaño del efecto, utilizando sus respectivas fórmulas. En este último caso, se siguió la clasificación de Cohen (1988, citado en Pardo y San Martín, 2015) para categorizar la intensidad del tamaño del efecto.

### Resultados

#### Análisis de la Plausibilidad de los Distractores

En primer lugar, se calculó la proporción de distractores plausibles para cada tipo de ítem, usando el criterio  $p_{nfd}$  como punto de corte, que se obtiene a través de la fórmula  $p_{nfd} = .10 - (p_j * .10)$ , y cuya lógica se puede consultar en Raymond et al. (2018). En los ítems de tres opciones, un 71.75% de los distractores eran plausibles, para los de cuatro, 62.08% y para los de cinco, 49.11%.

#### Efectos Posicionales de las Opciones de Respuesta sobre la Dificultad y Discriminación del Ítem

En segundo lugar, se evaluaron los efectos de la posición de la OC, el DMA y el DME en la dificultad y discriminación del ítem. Para ello se utilizó el Índice de Dificultad ( $p_j$ ) para la dificultad, y el Índice de Discriminación ( $D_j$ ) y la correlación ítem-resto del test para la discriminación. El Índice de Dificultad ( $p_j$ ) se calcula a partir de la proporción de aciertos en determinado ítem, el Índice de Discriminación ( $D_j$ ) obteniendo la diferencia de proporción de aciertos para el ítem entre el 25% de examinados con mayor puntuación y el 25% de menor puntuación en el test; por su parte, la correlación ítem-resto se calcula examinando la correlación de Pearson entre la puntuación en el ítem y en el resto del test, excluyendo el ítem en cuestión.

No se encontraron efectos de la posición de las tres opciones (OC, DME y DMA) en la discriminación del ítem (Índice de Discriminación y correlación ítem-resto), por lo que estos no se comentarán. Sí se obtuvieron, en cambio, efectos posicionales sobre la dificultad del ítem, que se presentarán a continuación.

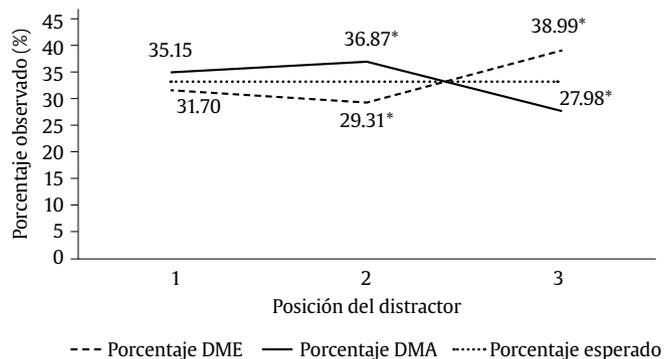
#### Efectos de la Posición de la Opción Correcta (OC)

La distribución de la OC fue uniforme para los ítems de tres y cuatro opciones, pero no para los de cinco,  $\chi^2(4) = 13,732, p = .008$ , encontrando esta opción en mayor proporción de lo esperado en

cuarto lugar y menor en primer lugar. No se encontraron efectos posicionales sobre la dificultad del ítem.

#### Efectos de la Posición del Distractor Menos Elegido (DME)

El DME no se distribuyó de forma uniforme en ninguno de los tres tipos de ítem. En los de tres opciones,  $\chi^2(2) = 11.509, p = .003$ , apareció con mayor frecuencia de la esperada en tercera posición y con menor de la esperada en segundo lugar (Figura 1). En ítems de cuatro opciones,  $\chi^2(3) = 24.334, p < .01$ , apareció con menor frecuencia de la esperada en primera posición, y con mayor de la esperada en la última posición (Figura 2). En los ítems de cinco opciones,  $\chi^2(4) = 22.945, p < .01$ , el DME apareció con menor frecuencia de la esperada en primera posición, y con mayor frecuencia en penúltima posición (Figura 3). En general, hubo una mayor probabilidad de lo esperado de que el DME se situase en las últimas posiciones y una menor probabilidad de encontrar esta opción en los primeros lugares.

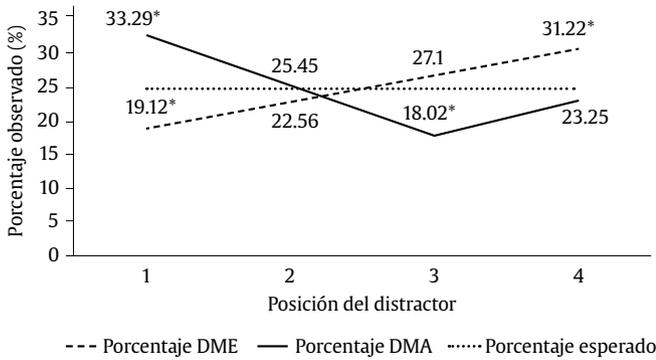


**Figura 1.** Distribución en Porcentaje del Distractor Menos Elegido (DME) y del Distractor más Atractivo (DMA), con respecto al Porcentaje Esperado, en Ítems de Tres Opciones de Respuesta.

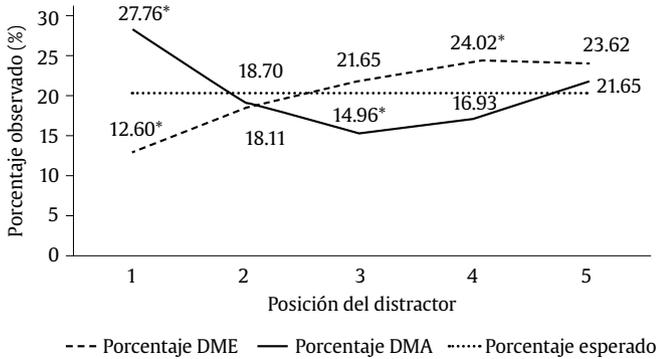
Nota. \*Diferencia significativa con respecto al porcentaje esperado, del 33.33%.

En cuanto a los efectos posicionales, la posición del DME tuvo efectos sobre la dificultad de los ítems de cuatro OOR,  $F(3, 723) = 6.970, p < .01; \omega^2 = .024$ . La prueba Tukey de comparaciones *post hoc* arrojó diferencias significativas de medias en la dificultad entre la primera y la cuarta posición ( $p = .001$ ) y entre la segunda y la cuarta posición ( $p = .001$ ). Al examinar las comparaciones de tendencias, se encontró una relación lineal entre la posición del DME y el Índice de Dificultad,  $F(1) = 18.371, p = .000$ , tal como aparece en la Figura 4. El tamaño del

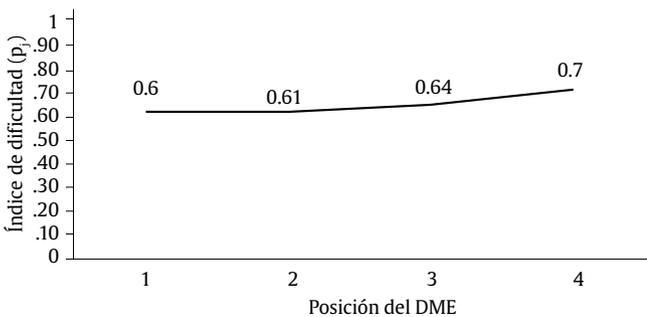
efecto fue pequeño.



**Figura 2.** Distribución en Porcentaje del Distractor Menos Elegido (DME) y del Distractor Más Atractivo (DMA), con respecto al Porcentaje Esperado, en Ítems de Cuatro Opciones de Respuesta.  
 Note. \*Diferencia significativa con respecto al porcentaje esperado, del 25%.

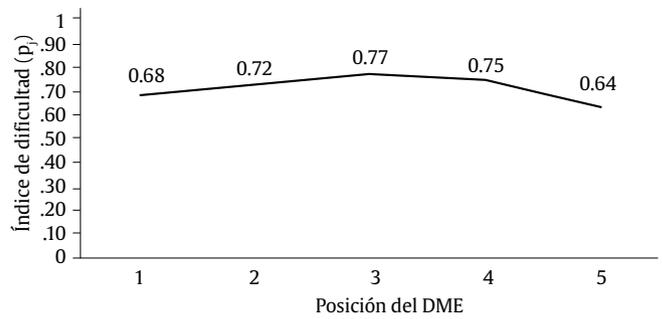


**Figura 3.** Distribución en Porcentaje del Distractor Menos Elegido (DME) y del Distractor Más Atractivo (DMA), con respecto al Porcentaje Esperado, en Ítems de Cinco Opciones de Respuesta.  
 Note. \*Diferencia significativa con respecto al porcentaje esperado, del 20%.



**Figura 4.** Gráfica del Efecto de la Posición del Distractor Menos Elegido (DME) Sobre el Índice de Dificultad en Ítems con Cuatro Opciones de Respuesta.

También aparecieron efectos posicionales sobre la dificultad en ítems de cinco OOR,  $F(4, 503) = 3.592, p = .007; \omega^2 = .033$ . La prueba de Tukey mostró diferencias de medias en el Índice de Dificultad entre las posiciones tercera y quinta ( $p = .001$ ), tal como se observa en la Figura 5. El tamaño del efecto fue también bajo en este caso.

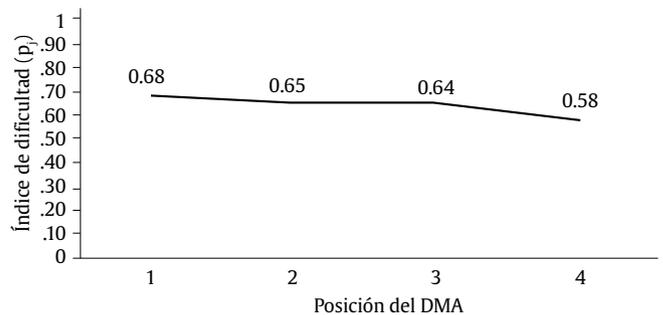


**Figura 5.** Gráfica del Efecto de la Posición del Distractor Menos Elegido (DME) Sobre el Índice de Dificultad en Ítems con Cinco Opciones de Respuesta.

**Efectos de la Posición del Distractor Más Atractivo (DMA)**

Al igual que en el caso del DME, la distribución del DMA no fue homogénea entre los distintos tipos de ítem. En los ítems de tres opciones,  $\chi^2(2) = 10.045, p = .007$ , este distractor apareció con mayor frecuencia de la esperada en segunda posición (Figura 1) y con menor frecuencia en la tercera. En ítems de cuatro opciones,  $\chi^2(3) = 35.096, p < .01$ , el DMA apareció más frecuentemente en primera posición y menos frecuentemente de lo esperado en tercer lugar (Figura 2). En los ítems de cinco opciones,  $\chi^2(4) = 25.248, p < .01$ , el DMA apareció con mayor frecuencia de la esperada en primera posición y con menor frecuencia en tercera posición (Figura 3).

Se halló un efecto de la posición del DMA en la dificultad de los ítems con cuatro OOR,  $F(3, 723) = 6.643, p < .01; \omega^2 = .023$ . La prueba Tukey mostró diferencias significativas entre las posiciones primera y cuarta ( $p = .000$ ) y se halló una relación lineal entre la posición del DMA y la dificultad del ítem,  $F(1) = 17.635, p = .000$ , tal como se observa en la Figura 6. El tamaño del efecto fue pequeño.



**Figura 6.** Gráfica del Efecto de la Posición del Distractor Más Atractivo (DMA) Sobre el Índice de Dificultad en Ítems de Cuatro Opciones de Respuesta.

**Discusión**

Un uso extendido de los exámenes de OM en ámbitos educativos trae consigo la necesidad de rigor y fiabilidad de estos instrumentos, razón por la cual vale la pena explorar fenómenos como el efecto de la posición de las opciones sobre los indicadores psicométricos, que si bien no genera variaciones drásticas en la puntuación, sí puede suponer diferencias relevantes en evaluaciones de importancia. En este estudio se examinan el efecto posicional de ítems de tres, cuatro y cinco OOR en los indicadores de la dificultad y discriminación, contribución singular del presente estudio. Los resultados indican cierto efecto de la posición de las OOR en la dificultad de los ítems, aunque no en la discriminación.

Distintos autores han encontrado sesgos de posición relacionados con la OC en la dirección de un efecto primacía, en la que los ítems se volverían más difíciles a medida que esta opción se coloca más hacia el final (Hagenmüller, 2020; Holzknrecht et al., 2020; Tellinghuisen

y Sulikowski, 2008), o de un sesgo central, en el que tanto examinadores como examinados prefieren posiciones centrales al colocar o seleccionar la OC (Attali y Bar-Hillel, 2003). Los datos aquí presentados no muestran ningún efecto de posición de la OC en la dificultad o la discriminación del ítem. No obstante, en ítems de cinco OOR se encontró una distribución sesgada de esta opción, lo cual podría indicar que a mayor número de opciones más difícil resulta equilibrar las posiciones de manera homogénea. Esto último podría suponer un problema si, por ejemplo, indujese un sesgo de posición en los examinados (Wevrick, 1962).

En cuanto a la posición del DME y del DMA, sí se observaron algunos efectos posicionales. La posición del DME mostró una correlación directa con la dificultad del ítem ( $p_i$ ) en ítems de cuatro OOR, por lo que cuanto más atrás se colocara el DME más fácil resultaría el ítem. Además, esta opción estuvo sobrerrepresentada en posiciones finales y la distribución no fue homogénea. Estos dos resultados podrían interpretarse en el sentido del “efecto anclaje” de Attali y Bar-Hillel (2003), según el cual los examinadores tienden a generar las OOR en orden de plausibilidad, de manera que de no redistribuir todas las opciones, las menos plausibles quedarían situadas al final con mayor probabilidad. Esta distribución propia del “efecto anclaje” se ha encontrado recientemente en el estudio de Lions et al. (2021), en el que los autores concluyen que los distractores se ordenan de mayor a menor plausibilidad. Una hipótesis posible en este caso sería que los examinados descartasen por defecto las últimas opciones, ya que se darían cuenta de que son generalmente menos atractivas, aumentando su probabilidad de acierto cuando el DME se encuentra al final (Lions et al., 2021).

En los ítems de cinco OOR la posición del DME también tuvo un efecto sobre la dificultad. Sin embargo, en este caso el ítem fue más fácil cuando esta opción se colocaba en la posición central. La distribución del DME también presentaba un patrón de “anclaje” en estos ítems, por lo que parece inusual que el ítem fuera más difícil si el DME se presentaba al final comparado con su presentación en la posición central. Una posible explicación, tal como sostienen Attali y Bar-Hillel (2003), sería que las posiciones centrales fueran más salientes, descartando los examinados el DME más fácilmente cuando se encuentra en esta posición (Lions et al., 2021).

En cuanto al DMA, en ítems de cuatro OOR la posición de este distractor estaba inversamente relacionada con la dificultad del ítem. Es decir, cuando el DMA se situaba en las últimas posiciones, el ítem resultaba más difícil. Este resultado contradice los resultados de Schroeder et al. (2012), que concluyeron que los DMA colocados en posiciones iniciales confundían en mayor medida a los examinados y mermaban la probabilidad de acierto. El “efecto anclaje” podría, por otro lado, aportar una explicación si los examinados esperaran las opciones más plausibles al principio, dudando más de su respuesta cuando estos se colocan al final. Más aún, la distribución del DMA tampoco fue homogénea, coincidiendo con esta disposición “anclada” de los distractores de más a menos plausible. No se analizó en este caso la posición relativa del DMA respecto a la OC, relevante en algunos estudios (Kiat et al., 2018; Shin et al., 2019).

En este trabajo no se encontraron efectos posicionales en la discriminación del ítem, medida con el Índice de Discriminación ( $D_i$ ) y con la correlación ítem-resto. Esto coincide con la literatura disponible, en la que se encuentran más efectos posicionales sobre la dificultad que sobre la discriminación del ítem. Sí existe, sin embargo, literatura que ha hallado algunos efectos posicionales sobre la discriminación (Attali y Bar-Hillel, 2003; Shin et al., 2019).

Los resultados de este trabajo tienen implicaciones sobre la construcción y utilización de ítems de OM; teniendo en cuenta este posible efecto posicional sobre los indicadores psicométricos se presentan aquí algunas recomendaciones. En primer lugar, se recomienda a los examinadores redistribuir todas las OOR, y no sólo la OC, siendo esta la práctica más común. A menudo la posición de los distractores se obvia, cuando según los resultados presentados, y en trabajos re-

lacionados (Cizek, 1994; Kiat et al., 2018; Shin et al., 2019), hay evidencia de un efecto de posición de los mismos. Esta práctica podría prevenir una distribución de “anclaje”, evitando que los distractores menos plausibles se coloquen frecuentemente al final, y con ella sesgos posicionales y la aplicación de estrategias de *testwiseness*.

Otra aplicación de estos resultados sería la mejora de ítems al borde del buen funcionamiento, a través de una reorganización de las OOR, lo cual podría variar sus indicadores psicométricos. Más allá de esto, la reorganización de las opciones podría servir a los examinadores para ajustar la dificultad del ítem mientras construyen el test (McNamara y Weitzman, 1945). Por ejemplo, si un ítem resultase poco discriminativo porque una gran mayoría de los examinados lo aciertan, se podría aumentar ligeramente la dificultad de forma artificial, colocando aquellos distractores más plausibles al final.

Por otra parte, se recomienda el uso preferente de ítems de tres OOR frente a los de cuatro o cinco. Algunas razones para ello son que estos ítems tienen mayor proporción de distractores plausibles, requieren menos tiempo de construcción y aplicación, permitiendo evaluar más conocimientos, facilitan la comprensión y sus indicadores psicométricos son de igual o mayor calidad que ítems de más opciones; es más, parecen ser menos sensibles a los efectos posicionales. En el presente estudio no ha habido efectos de posición para los ítems de tres OOR, mientras que los ítems de cuatro fueron los que más efecto de posición presentaron, coincidiendo con el trabajo de Holzknacht et al. (2020), cuyos resultados avalan un menor efecto de posición en ítems de tres OOR frente a los de cuatro. Adicionalmente, la redistribución de la OC y los distractores en ítems con menos opciones es más sencilla, compensando y previniendo así posibles sesgos.

Las limitaciones de este estudio se deben a la muestra empleada. La base de datos recogía una serie de exámenes ya aplicados, por lo que la manipulación experimental no fue posible, y tampoco habría sido ética, dado que se trata de evaluaciones reales con consecuencias en el desempeño académico; mientras tanto, la literatura muestra cómo en evaluaciones sin consecuencias, en las que sí podrían aplicarse manipulaciones experimentales éticamente, las tendencias de respuesta podrían variar, arriesgándose más los examinados, lo que podría aumentar el error aleatorio de medida. En consonancia con el diseño del presente estudio, sólo se pueden extraer conclusiones sobre las tendencias de respuesta de los examinadores y los examinados, y no sobre la causalidad de los efectos. Tampoco se ha podido supervisar el proceso de construcción o de aplicación de los test, por lo que esta pudo ser una variable extraña en el estudio, fuente de error aleatorio. Por otra parte, los exámenes procedían de evaluaciones de grado de la Universidad Autónoma de Madrid, cuyas conclusiones podrían no ser aplicables a otro tipo de contextos. No obstante, dado que los exámenes procedían de distintas disciplinas, de contextos de evaluación real, y fueron elaborados por distintos docentes, las conclusiones presentadas podrían generalizarse, pues sería representativo de lo que sucede en la realidad de la evaluación con OM. No siempre los desarrolladores de tests siguen fielmente las directrices establecidas, lo que compromete la calidad de la evaluación (Lions et al., 2022); al mismo tiempo, el efecto posicional fue lo suficientemente notable como para arrojar resultados significativos. Por otra parte, no se conocen estudios previos que hayan comparado los efectos de posición en ítems de tres, cuatro y cinco OOR, siendo también escasos los estudios que indagaran en el efecto de posición de los distractores menos plausibles (Kiat et al., 2018; Lions et al., 2021).

Futuras líneas de investigación podrían ahondar en los efectos de posición tanto de la OC como de los distractores sobre la dificultad y discriminación de los ítems de OM. La elaboración de diseños experimentales, contando con las limitaciones éticas y metodológicas de este ámbito, puede ayudar a esclarecer y unificar los resultados heterogéneos del efecto de posición. Este efecto de los distractores menos plausibles abre un nuevo curso de investigación. Paralelamente, el estudio de la posición de los DAPs también es una línea reciente que requiere de más estudios; en el presente trabajo se contradice el

“efecto primacía” de estos distractores, encontrado por autores como Schroeder et al. (2012), aunque en ambos casos los datos apuntan hacia diferencias en la puntuación no relacionadas con el conocimiento o habilidad que se pretende medir, esto es, hacia sesgos y posibles estrategias de *testwiseness*.

En resumen, los resultados obtenidos en el trabajo apuntan hacia un efecto de la posición de las OOR en la dificultad de los ítems de OM. No se hallaron efectos en relación a la posición de la OC, pero sí en cuanto a la posición de los distractores más y menos plausibles, líneas de reciente desarrollo. El tamaño del efecto fue pequeño, y aun así, la ubicuidad de este tipo de evaluaciones hace necesario el rigor y exactitud de la medida, prestando atención a fenómenos tales como los efectos posicionales.

### Extended Summary

The increasing popularity and implementation of multiple-choice (MC) assessment comes with the need for accurate evaluation. The advantages of this type of assessment would be its efficiency, the possibility to provide information about which contents have not been properly understood, as well as allowing to test a wider variety of topics (Gierl et al., 2017). Haladyna et al. (2013) have suggested guidelines for MC or multiple-choice test construction to ensure construct validity and adequate psychometric properties, as well as to avoid problems such as “testwiseness”, which is a special ability to guess the correct or key answer (KA) when examinees have limited or no knowledge of the topic evaluated. However, these construction guidelines are not widely implemented (Lions et al., 2022).

Option position was found to be a relevant factor in the accuracy of the measurements in multiple-choice evaluations. However, previous studies provide conflicting and scarce evidence about the effect of the arrangement of the response options on items' psychometric properties, that is, on their difficulty and discrimination. Regarding KA placement, some studies have found no significant effect on examinees scores (Cizek, 1994), while some have found a ‘primacy effect’ or a ‘middle bias’. Tellinghuisen and Sulikowski (2008) coined the term ‘primacy effect’, referring to the tendency to select the first option when presented with two equally attractive ones. This bias was also found in recent studies, such as Hagenmüller’s (2020). Schroeder et al. (2012) explain this effect by an increase in cognitive load as examinees search for the KA in later positions, causing confusion and fatigue. Holzknacht et al. (2020) found out via “eye tracking” technology that students spend significantly more time looking at the first response options compared to later ones. As for the ‘middle bias’, this effect was proposed by Attali and Bar-Hillel (2003) to reflect a tendency in both examiners and examinees to place and look for the KA in middle locations, and to avoid those at the beginning or the end (“edge aversion”).

Concerning distractor's position or incorrect answer's position, this has been a recent line of investigation in which studies mainly focus on the most plausible or attractive distractors. Findings suggest that highly attractive options at the beginning make items more difficult (Schroeder et al., 2012), especially when placed right before de KA (Kiat et al., 2018; Shin et al., 2019). According to Schroeder et al. (2012), highly plausible distractors placed at the first positions would confuse students and lower their success rate. Lions et al. (2021) found that higher plausibility distractors were often placed before the lower plausibility ones. This would happen, according to Attali and Bar-Hillel (2003), because developers recall better distractors easier and faster when constructing MC items, which these authors term an ‘anchoring effect’. The consequence of this would be that distractors would often be ordered in decreasing plausibility, which could tip the students on KA placement. The effect of lower plausibility or non-functioning distractors has yet to be explored, with few studies regarding its positional effects (Kiat et al., 2018).

An “anchored” distribution could mean that examinees expect lower plausibility options to be at the end, easily dismissing them, and increasing their chances of success (Lions et al., 2021).

In this study, response-option positional effects were assessed, using three, four, and five response-option items (ROI) from a sample of 1,989 items from university tests. There were no positional effects on item discrimination, but there were some effects on item difficulty.

In line with previous literature (Haladyna et al., 2019; Rodríguez, 2005; Tarrant et al., 2009), three-ROI presented the highest proportion of plausible distractors. While no effects of the KA position on the items' properties were found, distractor position seems to influence item difficulty. In four-ROI, when the most attractive distractor (the one with the most response choices) was placed in the last position, items became more difficult; meanwhile, when the least attractive distractor (the one with the least choices) was placed in the last position, items became easier. These two last results seem to be related to the ‘anchoring effect’ (Attali & Bar-Hillel, 2003), and the previously found tendency of examiners to write the distractors orderly with decreasing plausibility. Thus, students would expect the most plausible options in the first positions, and the least plausible ones in the last positions (Lions et al., 2021). With respect to the number of options, three-ROI appeared to be more robust to positional effects than four or five-ROI, as there were no effects of response option position on items with three options, and most positional effects were found on four-ROI.

These findings are relevant to MC assessment, and some recommendations could be made according to these results. Firstly, examiners should redistribute the position of all response options, and not just the KA, which is the usual practice (Gierl et al., 2017), as distractor position has proven to be an important and possibly confounding factor (Lions et al., 2021). This would help test developers avoid an ‘anchored’ distribution of options (Attali and Bar-Hillel, 2003). Secondly, three-ROI usage would be preferential, as this type of items show the highest proportion of plausible distractors, they require less time to construct and allow to test a wider range of topics, they have adequate psychometric properties compared to items with more options, and also seem more suitable when it comes to avoiding positional effects (Holzknacht et al., 2020). There were no positional effects found on three-ROI on this study, and four-ROI were the most prone to showing differences in item difficulty with regards to distractor position. Additionally, three response options would be easier to redistribute, preventing new positional biases.

Future studies should further investigate the effects of distractor-position on item difficulty and discrimination, which seems to be a novel and promising line of research. This study contributes to this field by comparing positional effects of response options on three, four, and five-ROI on item difficulty and discrimination, which, to our knowledge, has not been done before. Furthermore, the effect of higher and lower plausibility distractors was analyzed, which are novel and developing lines of research. According to these findings, positional effects on MC testing, although modest, should be taken into consideration, especially when examining is applied on high stakes evaluation.

### Conflicto de Intereses

Las autoras de este artículo declaran que no tienen ningún conflicto de intereses.

### Referencias

- Attali, Y. y Bar-Hillel, M. (2003). Guess where: The position of correct answers in multiple-choice test items as a psychometric variable. *Journal of Educational Measurement*, 40(2), 109-128. <https://doi.org/10.1111/j.1745-3984.2003.tb01099.x>
- Bresnack, A. E., Graves, P. E. y White, N. (1989). Multiple-choice testing: Question and response position. *The Journal of Economic Education*, 20(3), 239-245. <https://doi.org/10.2307/1182299>

- Brooks, G. P. y Johanson, G. A. (2003). TAP: Test analysis program. *Applied Psychological Measurement*, 27(4), 303-304. <https://doi.org/10.1177/0146621603027004007>
- Carnegie, J. A. (2017). Does correct answer distribution influence student choices when writing multiple choice examinations? *The Canadian Journal for the Scholarship of Teaching and Learning*, 8(1), 1-13. <https://doi.org/10.5206/cjsotl-rcacea.2017.1.11>
- Cizek, G. J. (1994). The effect of altering the position of options in a multiple-choice examination. *Educational and Psychological Measurement*, 54(1), 8-20. <https://doi.org/10.1177/0013164494054001002>
- Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, 10(1), 3-31. <https://doi.org/10.1177/001316445001000101>
- Gierl, M. J., Bulut, O., Guo, Q. y Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*, 87(6), 1082-1116. <https://doi.org/10.3102/0034654317726529>
- Hagenmüller, B. (2020). On the impact of the response options' position on item difficulty in multiple-choice-items. *European Journal of Psychological Assessment*, 37(4), 1-10. <https://doi.org/10.1027/1015-5759/a000615>
- Haladyna, T. M. y Rodríguez, M. C. (2013). *Developing and validating test items*. Routledge.
- Haladyna, T. M., Rodríguez, M. C. y Stevens, C. (2019). Are multiple-choice items too fat? *Applied Measurement in Education*, 32(4), 350-364. <https://doi.org/10.1080/08957347.2019.1660348>
- Holzknicht, F., McCray, G., Eberharter, K., Kremmel, B., Zehentner, M., Spiby, R. y Dunlea, J. (2020). The effect of response order on candidate viewing behaviour and item difficulty in a multiple-choice listening test. *Language Testing*, 38(1), 41-61. <https://doi.org/10.1177/0265532220917316>
- Kiat, J. E., Ong, A. R. y Ganesan, A. (2018). The influence of distractor strength and response order on MCQ responding. *Educational Psychology*, 38(3), 368-380. <https://doi.org/10.1080/01443410.2017.1349877>
- Lions, S., Monsalve, C., Dartnell, P., Blanco, M. P., Ortega, G. y Lemarié, J. (2022). Does the response options placement provide clues to the correct answers in multiple-choice tests? A systematic review. *Applied Measurement in Education*, 35(2), 133-152. <https://doi.org/10.1080/08957347.2022.2067539>
- Lions S., Monsalve C., Dartnell P., Godoy M. I., Córdova N., Jiménez D., Blanco M. P., Ortega G. y Lemarié J. (2021). The position of distractors in multiple-choice test items: The strongest precede the weakest. *Frontiers in Education*, 6, Artículo 731763. <https://doi.org/10.3389/fe-duc.2021.731763>
- Marcus, A. (1963). The effect of correct response location on the difficulty level of multiple-choice questions. *Journal of Applied Psychology*, 47(1), 48-51. <https://doi.org/10.1037/h0042018>
- McNamara, W. J. y Weitzman, E. (1945). The effect of choice placement on the difficulty of multiple-choice questions. *Journal of Educational Psychology*, 36(2), 103-113. <https://doi.org/10.1037/h0060835>
- Raymond, M. R., Stevens, C. y Bucak, S. D. (2018). The optimal number of options for multiple-choice questions on high-stakes tests: Application of a revised index for detecting nonfunctional distractors. *Advances in Health Sciences Education*, 24(10), 141-150. <https://doi.org/10.1007/s10459-018-9855-9>
- Rodríguez, 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. <https://doi.org/10.1111/j.1745-3992.2005.00006.x>
- Pardo, A. y San Martín, R. (2015). *Análisis de datos en ciencias sociales y de la salud II*. Síntesis.
- Schroeder, J., Murphy, K. L. y Holme, T. A. (2012). Investigating factors that influence item performance on ACS exams. *Journal of Chemical Education*, 89(3), 346-350. <https://doi.org/10.1021/ed101175f>
- Shin, J., Bulut, O. y Gierl, M. J. (2019). The effect of the most-attractive-distractor location on multiple-choice item difficulty. *The Journal of Experimental Education*, 88(4), 643-659. <https://doi.org/10.1080/00220973.2019.1629577>
- Tarrant, M., Ware, J. y Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. *BMC Medical Education*, 9(1), Artículo 40. <https://doi.org/10.1186/1472-6920-9-40>
- Tellinghuisen, J. y Sulikowski, M. M. (2008). Does the answer order matter on multiple-choice exams? *Journal of Chemical Education*, 85(4), 572-575. <https://doi.org/10.1021/ed085p572>
- Wevrick, L. (1962). Response set in a multiple-choice test. *Educational and Psychological Measurement*, 22(3), 533-538. <https://doi.org/10.1177/001316446202200310>
- Wilbur, P. H. (1966). *Positional response set in the multiple choice examination* (tesis de doctorado). University of Southern California Digital Library. <https://doi.org/10.25549/usctheses-c18-114359>