

El Modelo Multiple Choice y su utilidad para reducir distractores

Facundo Juan Pablo Abal - Universidad de Buenos Aires, Argentina

Gabriela Susana Lozzia - Universidad de Buenos Aires, Argentina

María Silvia Galibert - Universidad de Buenos Aires, Argentina

María Ester Aguerra - Universidad de Buenos Aires, Argentina

Horacio Félix Attorresi - Universidad de Buenos Aires, Argentina

Resumo

La reducción de distractores puede resultar una herramienta útil para mejorar la calidad psicométrica de una prueba o acortarla sin que se perjudique la validez de contenido. El objetivo de este trabajo es presentar criterios para la reducción de distractores surgidos del análisis de las curvas características de un modelo de la Teoría de Respuesta al Ítem: el Multiple Choice de Thissen y Steinberg (1997). Se categorizaron posibles comportamientos inadecuados de los distractores en dos grupos: curva creciente y curva horizontal. Estas categorías orientan respecto de qué alternativa conviene eliminar en función del grado de inadecuación de funcionamiento del distractor. Este procedimiento permite conservar el conjunto más homogéneo y más plausible de opciones. Se discuten las ventajas de implementar estas pautas en distintos momentos de la construcción o adaptación de una prueba.

Palabras clave: análisis de distractores; análisis de ítems; Teoría de Respuesta al Ítem; Modelo Multiple Choice; Modelo de Thissen y Steinberg.

The Multiple Choice Model and its usefulness to reduce distractors

Abstract

The reduction in the number of distractors may be a useful tool to improve the psychometric quality of a test or to shorten it without affecting content validity. The purpose of this study is to put forward criteria for the reduction of distractors as suggested in the characteristic curves analysis of an Item Response Theory model: Thissen and Steinberg's Multiple Choice (1997). Possible unsuitable distractor behaviors were classified in two groups: growing curve and horizontal curve. These categories allow to make a decision on which alternative to remove based on the degree of unsuitability of the distractor. This procedure leads to a more homogeneous whole with more plausible options. It is discussed which advantages there are in the implementation of these guidelines at different moments in the test construction or adaptation.

Keywords: distractor analysis; item analysis; Item Response Theory; Multiple Choice Model; Thissen and Steinberg's Model.

O Modelo Múltipla Escolha e sua utilidade para a redução dos distratores

Resumo

A redução dos distratores pode ser uma ferramenta útil para melhorar a qualidade psicométrica de um teste ou reduzi-lo sem que a validade de conteúdo seja prejudicada. O objetivo deste trabalho é apresentar critérios para a redução dos distratores surgidos da análise das curvas características de um modelo da Teoria de Resposta ao Item: A Múltipla Escolha de Thissen e Steinberg (1997). Foram registrados possíveis comportamentos inadequados dos distratores em dois grupos: curva crescente e curva horizontal. Estas categorias permitem decidir qual a alternativa que é possível eliminar, baseado no grau de inadequação do funcionamento do distrator. Este procedimento permite ter um conjunto de opções mais homogêneo e mais plausível. Se discute as vantagens de fazer implementações destes critérios em diferentes etapas de elaboração e adaptação de um teste.

Palavras chave: Análise de distratores; análise de itens; teoria de desposta ao item; modelo Multiple Choice; modelo de Thissen y Steinberg.

Durante el proceso de construcción o adaptación de una prueba estructurada es posible que se presente la necesidad de reducir la cantidad de distractores de los ítems de elección múltiple. Son variadas las situaciones en las que puede instrumentarse esta estrategia de

modificación, las más comunes suelen ser: a) presencia de inadecuadas propiedades psicométricas de la prueba generadas por fallas en la homogeneidad y plausibilidad de los distractores, b) disminución del tiempo disponible para realizar la administración, c) necesidad de incluir

Endereço de correspondência:

Dirección Postal: Zuviría 5691, CP: 1439, Ciudad de Buenos Aires, Teléfono/Fax: (5411) 4637-0923, E-mail: fabal@psi.uba.ar

más ítems en la evaluación sin que la prueba se torne excesivamente extensa, o d) adaptación de los ítems a las características especiales de la población a evaluar (por ejemplo, la adaptación de un instrumento validado para jóvenes a una población infantil podría considerar la reducción de alternativas dado que es menor la capacidad de los niños para manejar gran cantidad de opciones).

La decisión respecto de cuál es el distractor que convendría eliminar en cada ítem debería considerarse con las mismas directrices que guían la redacción de los reactivos (Haladyna, Downing & Rodríguez, 2002; Martínez, Moreno & Muñiz, 2005; Osterlind, 1989). Estas enfatizan la necesidad de construir distractores plausibles y homogéneos en contenido. Sin embargo, estas propiedades dependen del número de opciones que presenta el ítem. A mayor cantidad de alternativas aumenta el riesgo de construir distractores irrelevantes y, al mismo tiempo, más heterogéneos en contenido. Por ende, reducir el número de alternativas posibilita obtener distractores más plausibles y facilita la homogeneidad en el contenido de los mismos.

Más allá de considerar la plausibilidad y homogeneidad como criterios rectores, la tarea de decidir cuál distractor eliminar puede resultar compleja si no se dispone de herramientas psicométricas que evalúen la eficacia de las alternativas incorrectas. Desde la perspectiva de la Teoría Clásica de Test (TCT), el análisis de los distractores se efectúa mediante indicadores que comparan el desempeño de los individuos que obtuvieron los puntajes inferiores y superiores en el test. Por ejemplo, se recomienda cotejar las distribuciones de frecuencias (Aiken, 2003; Cohen & Swerdlik, 2001) o la media aritmética de cada alternativa (Martínez-Airas, 1995). Más recientemente, Martínez-Arias, Hernández-Lloreda y Hernández-Lloreda (2006) sugirieron la representación tabular o gráfica de las frecuencias de elección de las diferentes alternativas para cada uno de los cuartiles o quintiles de la distribución de la habilidad. Otro posible análisis es la correlación biserial de los distractores con el puntaje total del test (Martínez-Arias, 1995). A su vez, se propuso estudiar la homogeneidad de las proporciones para decidir acerca de la equiprobabilidad de elegir a las alternativas incorrectas (Muñiz, 1994).

La Teoría de Respuesta al Ítem (TRI) ha generado importantes contribuciones para profundizar y enriquecer el análisis de los distractores. En el marco de esta teoría se han propuesto modelos para explicar

las respuestas a ítems de elección múltiple tanto con aproximaciones paramétricas (Bock, 1972, 1997, Samejima, 1979, Thissen y Steinberg, 1997) como no-paramétricas (Ramsay, 1997). Específicamente, el Modelo Multiple Choice (MMC) desarrollado en los trabajos de Thissen, Steinberg y Fitzpatrick (1989) y de Thissen y Steinberg (1997) permite trazar en un mismo gráfico las curvas de probabilidad de respuesta de la opción correcta y de los distractores de un ítem de elección múltiple en función del nivel de habilidad del individuo. Esto hace posible estudiar no sólo cuántos eligieron cada alternativa sino también qué clase de sujetos, según su nivel de habilidad, se inclinaron por cada opción.

El objetivo de este trabajo es presentar las directrices establecidas por los autores para la reducción de distractores a partir del análisis de las curvas del MMC de Thissen y Steinberg. La aplicación de estas pautas contribuye a maximizar la homogeneidad y plausibilidad de los distractores utilizando una herramienta específica.

Desarrollo

Características del Modelo Multiple Choice de Thissen y Steinberg

Cuando un sujeto elige una de las alternativas de un ítem de elección múltiple se espera que lo haga motivado por una razón pertinente a la habilidad que mide. Dado un nivel de dificultad, si es más habilidoso se supone que elegirá la clave y si es menos competente es probable que sea atraído por alguno de los distractores. Pero la elección de una opción también puede efectuarse por una razón ajena a la habilidad; ya sea por puro azar frente a la ignorancia de lo que se debe responder, ya sea por algún razonamiento falaz o porque la opción resulte atractiva por factores no pertinentes al rasgo que se desea medir. Thissen y Steinberg (1997) denominaron a esta respuesta como *sin saber (Don't Know)* y la diferenciaron de la respuesta azarosa en tanto es un concepto más general, ya que la probabilidad de elegir cada una de las alternativas no es necesariamente homogénea. Esto es, el individuo que carece de un criterio pertinente para identificar la respuesta correcta puede ser más atraído por una alternativa que por otra.

Basados en los desarrollos de Bock (1972, 1997) y Samejima (1979), Thissen y Steinberg (1997) propusieron un modelo para el análisis del comportamiento

de ítems de respuesta múltiple. Este modelo establece una relación funcional entre el nivel de rasgo latente del individuo (θ) y la probabilidad de contestar a la alternativa h . Dado que h puede adoptar tantos valores como alternativas de respuestas presente el ítem ($h = 1, 2, \dots, m$), la misma función puede aplicarse para describir el funcionamiento de la clave y los distractores.

$$P(x = h / \theta) = \frac{\exp(a_h \theta + c_h)}{\sum_{k=0}^m \exp(a_k \theta + c_k)} + d_h \frac{\exp(a_0 \theta + c_0)}{\sum_{k=0}^m \exp(a_k \theta + c_k)}$$

También el modelo incluye la respuesta *sin saber* como una categoría hipotética denominada *categoría 0* o *DK*. De este modo, más allá de que el individuo puede optar efectivamente entre m categorías, el modelo contempla la posibilidad de pertenecer a $m+1$ categorías ($k = 0, 1, 2, \dots, m$). La expresión del modelo que incluye la probabilidad asignada a la categoría DK es:

$$P(x = 0 / \theta) = \frac{\exp(a_0 \theta + c_0)}{\sum_{k=0}^m \exp(a_k \theta + c_k)}$$

El modelo descompone la probabilidad de elegir la alternativa h , para cada valor dado de habilidad θ , $P(x = h / \theta)$, como la suma de probabilidades de dos sucesos mutuamente excluyentes: “elegir h y ser de la clase de los que eligen con criterio” (primer término) o “elegir h y ser de la clase de los que no saben” (segundo término). A su vez, la probabilidad de este segundo suceso se descompone como el producto de la probabilidad condicional de “elegir la alternativa h dado que se contestó sin saber” (d_h), por la probabilidad de pertenecer a la clase de los que no saben (*categoría 0* o *DK*). En otras palabras lo que el modelo expresa es que, para cada nivel de habilidad, la proporción de sujetos que eligieron la alternativa h puede desglosarse en la proporción de los que la eligieron por razones pertinentes más la proporción de los que la eligieron por otras razones; esta última deviene del porcentaje de sujetos que eligió esta opción entre los que no saben ponderada por el peso de la categoría *DK* en el total de sujetos.

La estimación de los parámetros se realiza por máxima verosimilitud marginal a través del programa MULTILOG™ (Thissen, 1991) y según ciertas restricciones (para más detalles ver Thissen & Steinberg, 1997):

$$\sum_{k=0}^m a_k = \sum_{k=0}^m c_k = 0 \quad \text{y} \quad \sum_{h=1}^m d_h = 1$$

Los parámetros c reflejan la frecuencia relativa de elección de cada alternativa y los parámetros a se relacionan con el crecimiento y monotonía de las funciones. Por lo general, el mayor valor de los parámetros a suele ser el correspondiente a la clave en tanto que el de *DK* está asociado con el menor valor. Valores intermedios de a producen funciones no monótonas que corresponden a los distractores (Thissen & Steinberg, 1997).

Una de las dificultades del modelo es la gran cantidad de parámetros por estimar que involucra (Abad, 2001). Sería conveniente una reducción de los mismos imponiendo restricciones convenientes que requerirían de algún previo conocimiento o hipótesis respecto del comportamiento de las curvas. Con todo, Thissen y Steinberg (1997) señalaron que la información extraída de la aplicación del modelo es interpretable aún cuando el ajuste o las restricciones no sean exhaustivos. Esto se convierte en una gran ventaja dado que es posible su utilización en las etapas iniciales del análisis cuando la calidad psicométrica de la escala en construcción puede no ser del todo adecuada. Incluso, la interpretación de las curvas de probabilidad de las alternativas puede constituirse como una fuente de información muy importante para obtener mejores indicadores de unidimensionalidad del constructo y fiabilidad del instrumento.

Según Thissen y Steinberg (1997) un ítem tiene un comportamiento óptimo cuando la Curva de la Clave (CC) es monótona creciente y asintótica a 1 (máxima probabilidad) en los niveles de habilidad más altos. Es decir, cuanto más hábiles son los individuos mayor probabilidad tendrán de contestar correctamente el ítem. Consecuentemente, a mayor nivel de habilidad también se espera una menor probabilidad de elección de respuesta *sin saber* lo que implicaría, para un adecuado funcionamiento, que la curva *DK* sea decreciente y asintótica a 0 en los niveles más altos de la habilidad. Con respecto a las curvas de los distractores, se considera un buen comportamiento si, al igual que *DK*, son asintóticas a 0 hacia los niveles más altos de la habilidad dado que esto implicaría que los sujetos más habilidosos fueron atraídos por la opción correcta. No obstante la monotonía no es en general una característica esperable en estas curvas. Por ejemplo en ítems de habilidades puede haber una

jerarquía conceptual de los errores que produce curvas no monótonas con máximos en diferentes niveles de habilidad coherentes con el orden de los errores conceptuales. De esta forma, podría asociarse el aspecto erróneo que refleja el contenido del distractor con un rango de habilidad determinado.

Criterios para la Eliminación de Distractores

Tomando como referencia la configuración ideal de las curvas se categorizaron los comportamientos inadecuados de los distractores en dos grupos. Uno de ellos considera la presencia de un distractor con curva creciente hacia los niveles altos de habilidad. El otro grupo contempla una curva de distractor relativamente horizontal. Estas categorías reflejan diferentes tipos de defectos y distintos grados de inadecuación del funcionamiento del distractor. Esto repercute al momento de decir cuál es la alternativa incorrecta que conviene eliminar para maximizar la calidad psicométrica del ítem. La curva DK no ha sido empleada para establecer los criterios de exclusión de los distractores ya que no se consideró fundamental la información que aporta para esta toma de decisión.

Para ilustrar la aplicación de estos criterios se exhiben dos gráficas de hipotéticos ítems de elección múltiple (ver Figura 1 y 2). Cada figura incluye las curvas de la Clave, del Distractor que se debería Eliminar (DE), de los demás distractores (D1 y D2) y la de la categoría 0 (DK). En la Tabla 1 se pueden apreciar los parámetros utilizados para graficar las curvas de cada ítem.

1. Presencia de distractor con curva de probabilidad creciente. Esta configuración se presenta cuando una proporción relativamente grande de los individuos más habilidosos resultan atraídos por una opción incorrecta (ver las curvas Clave y DE en la Figura 1). Muñiz (1994) aseguró que no es infrecuente encontrar distractores que resulten más atractivos para los individuos de niveles altos de habilidad que para los menos habilidosos. Este autor sostiene que este tipo de funcionamiento defectuoso se explica por la presencia de un distractor que contiene información de un nivel elevado. Este contenido problematiza a los más habilidosos en tanto que es inadvertido por los sujetos de niveles más bajos.

La presencia de un distractor con curva creciente es un indicio de que un factor ajeno a la habilidad que se pretende evaluar ha incidido en las respuestas de los

individuos. Esta alternativa afecta negativamente en la calidad del reactivo dado que el crecimiento de la curva del distractor puede o bien ir acompañado por el decrecimiento de la respuesta correcta para valores altos de la habilidad o bien impedir que la misma alcance la máxima probabilidad de elección. A su vez, también reduce la efectividad de las otras alternativas incorrectas en virtud del atractivo despertado por este distractor inadecuado. La consecuencia de este funcionamiento, a nivel global del test, se traduce en una dificultad para alcanzar una aceptable unidimensionalidad, es decir atenta contra ésta.

Lo que deja en evidencia esta configuración de curvas es que uno de los distractores del ítem no es homogéneo respecto de los otros. Los individuos más habilidosos están reaccionando ante el ítem de manera ligeramente diferente de lo que esperaba el autor de la prueba. El MMC orienta respecto de qué distractor podría ser la fuente del malentendido. Quedará en manos del autor de la prueba encontrar una justificación que explique el comportamiento defectuoso de la alternativa incorrecta en cuestión. Esta razón puede encontrarse tanto en la apariencia o en el contenido del distractor como también en la relación que éste tiene con la clave. Un contenido ambiguo de la respuesta correcta puede hacer que un distractor altamente plausible se muestre como más atractivo. Como sugirieron Cohen y Swerdlik (2001) entrevistar a los sujetos habilidosos que optaron por esta alternativa incorrecta brindaría más información cualitativa para explicar el comportamiento inadecuado del ítem.

Un caso extremo de funcionamiento inadecuado de ítem perteneciente a esta categoría se observa cuando un distractor ha funcionado como debía haberlo hecho la clave. Es decir, cuando la curva del distractor no sólo es creciente sino que supera la probabilidad de elección de la clave en los valores más elevados de la habilidad. Esta configuración puede resultar frecuente en reactivos de elevada dificultad con presencia de un distractor muy atractivo respecto de otros irrelevantes. En términos de Osterlind (1989) se trata de una alternativa excesivamente plausible (*overly plausible*). La presencia de este distractor también va en detrimento de la calidad del ítem y del test dado que perjudica el desempeño de los individuos más competentes.

2. Presencia de distractor con curva de probabilidad relativamente horizontal. Si al analizar el comportamiento de la respuesta correcta de un ítem

se observa que los individuos con puntaje altos y bajo en el test la eligieron en similar proporción se afirma que el reactivo tiene poca capacidad de discriminación. Análoga expresión es utilizada cuando un distractor no logra diferenciar a sujetos poco hábiles de los más competentes (ver en Figura 2 que la curva DE se mantiene constante en torno a una probabilidad de .20). En el MMC, la inexistencia de pendiente en una curva refleja que la probabilidad de elegir esa alternativa es aproximadamente la misma cualquiera sea el nivel de habilidad del sujeto. Como se mencionó anteriormente, es esperable que cada distractor proporcione información relevante respecto del error más común que puede cometer un sujeto según su nivel de habilidad. Por el contrario, un distractor sin capacidad de discriminación demuestra que el contenido del mismo no es esencial para determinar el nivel de la habilidad que se pretende medir.

Un caso particular de esta categoría se da ante la presencia de un distractor con escasa o nula probabilidad de elección a lo largo de todos los niveles de habilidad, lo cual refleja poca plausibilidad. Desde la perspectiva de la TCT, suele recomendarse la revisión de la alternativa incorrecta menos atractiva. La diferencia radica en que el MMC permite corroborar si la escasa elección de esta alternativa se da en todo el recorrido de la habilidad o se acumula en algún nivel de esta variable. En el caso de registrarse esta última situación indicaría que el distractor es útil para estudiar el patrón de respuesta de los individuos que corresponden a ese nivel de habilidad.

Se supone que la reducción de esta alternativa no alterará significativamente el comportamiento de las otras curvas. Por esto mismo, este criterio debería adoptarse para los ítems que han tenido un funcionamiento óptimo pero que igual requieren de una reducción de distractores. Si el ítem con comportamiento adecuado no presenta ningún distractor con curva horizontal y de baja probabilidad de elección entonces, en estos casos, habría que eliminar la opción menos atractiva como lo indica la TCT.

Discusión

El avance de los modelos de la TRI ha permitido profundizar el análisis de los ítems para maximizar la calidad del test. A través del MMC se pueden obtener indicadores eficaces para evaluar el comportamiento no sólo de la respuesta correcta sino también de sus

distractores. De esta forma, el modelo brinda orientaciones respecto de cuál es la decisión más acertada al momento de elegir un distractor para eliminar.

Si bien se podría suponer que la descripción gráfica del comportamiento del reactivo provista por el MMC es similar al análisis de distractores clásico explicado por Martínez-Arias, Hernández-Lloreda y Hernández-Lloreda (2006) las diferencias son notables. En principio, el MMC no depende del puntaje total bruto sino que se basa en una estimación del nivel de rasgo latente. La precisión alcanzada con este modelo de la TRI es superior en tanto que las curvas no se construyen únicamente con los cuartiles de la distribución de la habilidad sino que se representan en función de todos los niveles estimados del rasgo. A su vez, el MMC incluye el análisis de la respuesta de aquellos individuos que responden *sin saber*, aspecto que es difícilmente abordable desde la TCT. Esto se suma a las múltiples ventajas que de por sí presenta analizar los ítems desde una perspectiva más rigurosa como la TRI.

La riqueza del análisis de las curvas del MMC permite pensar una amplia gama de modificaciones para los ítems con comportamiento inadecuado. En particular, este análisis ha hecho énfasis en el funcionamiento de los distractores. No obstante, debe recordarse que cada ítem es un sistema complejo que puede ser abordado desde cualquiera de sus variables. Por ejemplo, podrían establecerse criterios de evaluación de los ítems que consideren sólo la forma de la CC u otros más abarcadores que incorporen al análisis la información que suministra la curva DK. Así también, cabe aclarar que si bien los criterios presentados en este artículo fueron definidos en función de encontrar un método que oriente la reducción de distractores también pueden ser aplicados para detectar las alternativas por modificar o reemplazar si fuere necesario.

Recapitulando, para determinar qué alternativa incorrecta conviene eliminar, el constructor debe resaltar que la eficacia de un distractor radica en su capacidad para atraer a sujetos poco hábiles y ser descartado por sujetos con mayor capacidad. Cualquier funcionamiento que se aleje de éste resultará inadecuado y perjudicará la calidad psicométrica de la escala:

1. Si un distractor resulta más atractivo para los individuos más hábiles que para los menos competentes su curva de probabilidad resultará creciente. En este caso, se debería aplicar el criterio 1.
2. Si un distractor resulta igual de atractivo para los individuos más competentes y los menos hábiles

su curva de probabilidad resultará horizontal. Entonces se debería aplicar el criterio 2.

3. Si un distractor no resulta atractivo ni para los individuos más hábiles ni para los inhábiles la curva se mostrará plana con un nivel de probabilidad cercano a cero. Esta situación ha sido considerada como un caso particular del criterio 2.

La aplicación de los criterios de reducción de distractores desarrollados en el presente artículo debe ser contemplada en función de los objetivos del investigador y de la etapa en que se encuentre la construcción del instrumento:

1. Si el objetivo es mejorar la calidad psicométrica de una prueba, la reducción de distractores le permitirá conservar las alternativas más homogéneas en contenidos. Estudiar esta propiedad puede ser necesario en ítems de elección múltiple de pruebas del tipo *best-answer* (Osterlind, 1989) donde se torna difícil determinar a priori la homogeneidad de las alternativas. Ya desde una perspectiva clásica se contempla que la modificación del número de opciones podría tener efectos distintos sobre la eficacia del ítem según el nivel de habilidad de los sujetos (Muñiz, 1994). Es probable que los individuos menos habilidosos, al ser "seducidos" por distractores más plausibles, obtengan un peor resultado que si contestasen por azar. Como se ha mencionado, el hecho de que un distractor presente una curva creciente podría representar un indicio de que la resolución del ítem compromete otras variables. Por lo tanto, dejar de lado tal distractor posiblemente favorezca a la determinación de la unidimensionalidad del constructo.
2. Si el investigador pretende disminuir el tiempo de administración, la reducción de distractores es una modificación posible cuando no se pueden eliminar ítems para preservar la validez de contenido. Si la prueba contiene reactivos con óptima calidad psicométrica deberá adoptar principalmente el criterio que contempla la eliminación de la alternativa con curva relativamente horizontal con baja probabilidad de elección. Evidentemente, en esta situación también se encuentra el investigador que precisa incluir más ítems en la evaluación sin modificar significativamente su extensión y quien necesita adaptar la cantidad de alternativas por particularidades de la población estudiada.
3. Si el constructor se encuentra en la etapa de redacción de ítems sería recomendable que incorpore

más distractores de los necesarios a sabiendas de que podrá optar, mediante estos criterios, por aquellos que resulten de mayor utilidad para los fines del test. Es decir, así como es habitual depurar los ítems con funcionamiento defectuoso, se podría realizar un análisis previo de los distractores inadecuados para descartarlos. Esto evitaría la redacción de nuevos ítems y la repetición de ensayos en pruebas pilotos y jueces expertos. Ahora bien, al implementar la reducción de distractores el constructor también debe considerar que más allá de que la mayoría de las pruebas utilicen cuatro o cinco alternativas, numerosas investigaciones han demostrado que tres es la cantidad óptima de opciones para las pruebas de elección múltiples (e.g. Abad, Olea & Ponsoda, 2001; Haladyna, Downing & Rodríguez, 2002; Rogers & Harley, 1999). Estos estudios sirven al investigador para saber cuántos distractores puede eliminar sin correr el riesgo de que aumente la respuesta azarosa. No obstante, la aplicación del MMC con este fin debe realizarse de forma estratégica dado que, como todos los modelos de la TRI, requiere de un tamaño muestral considerable que encarece los costos de una prueba piloto.

La sistematización y jerarquización de los posibles defectos de los distractores son completamente generalizables a todas las pruebas de elección múltiple sobre las que se aplique el modelo de Thissen y Steinberg. La interpretación de los resultados que proporciona el Modelo Multiple Choice es dificultosa por la gran cantidad de variables que incluye. Por tal motivo, se espera que la determinación de estos criterios implique un aporte metodológico que facilite la tarea de quienes utilicen este modelo y aborden el análisis de distractores.

Referencias

- Abad, F. J. (2001). *Algunas soluciones para la estimación del modelo de elección múltiple de Thissen y Steinberg*. Tesis doctoral. Madrid: UAM.
- Abad, F. J., Olea, J. & Ponsoda, V. (2001). Analysis of the optimum number of alternatives from the item response theory. *Psicothema*, 13, 152-158.
- Aiken, L. R. (2003). *Tests Psicológicos y Evaluación*. Undécima Edición. México: Pearson.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more latent categories. *Psychometrika*, 37, 29-51.

- Bock, R. D. (1997). The Nominal Categories Model. En W. J. van der Linden. & R. K. Hambleton (Eds.). *Handbook of Modern Item Response Theory* (pp 33-49). New York: Springer.
- Cohen, R. J. & Swerdlik, M. E. (2001). *Pruebas y evaluación psicológicas*. México: McGraw-Hill.
- Haladyna, T. M., Downing, S. M. & Rodríguez, M. C. (2002). A review of multiple-choice item-writing guidelines. *Applied Measurement in Education*, 15, 309-334.
- Martínez, R., Moreno, R. & Muñoz, J. (2005). Construcción de los ítems. En J. Muñoz, A. M. Fidalgo, E. García-Cueto; R. Martínez & R. Moreno. *Análisis de los ítems* (pp. 9 - 52). Madrid: La Muralla.
- Martínez-Arias, M. R. (1995). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid: Síntesis.
- Martínez-Arias, M. R., Hernández-Lloreda, M. V. & Hernández-Lloreda, M. J. (2006). *Psicometría*. Madrid: Alianza Editorial.
- Muñoz, J. (1994). *Teoría Clásica de Test*. Madrid: Pirámide.
- Osterlind, S. J. (1989). *Constructing Test Items*. Boston: Kluder Academic Publishers.
- Ramsay, J. O. (1997). A Functional Approach to Modeling Test Data. En W. J. van der Linden. & R. K. Hambleton (Eds.). *Handbook of Modern Item Response Theory*. (pp 381-394). New York, Springer.
- Rogers W. T. & Harley, D (1999). An empirical comparison of three-and-four-choice ítems and test: Susceptibility to testwiseness and internal consistency reliability. *Educational and Psychological Measurement*, 59, 234-247.
- Samejima, F. (1979). A New Family of Models for the Multiple Choice Item (Research Report #79-4). Knoxville, TN: University of Tennessee, Department of Psychology.
- Thissen, D. (1991). *MULTILOG™. User's Guide. Multiple, Categorical Item analysis and Test Scoring Using Item Response Theory*. Scientific Software. Inc.
- Thissen, D. & Steinberg, L. (1997). A Response Model for Multiple-Choice Items. En W. J. van der Linden. & R. K. Hambleton (Eds.). *Handbook of Modern Item Response Theory* (pp 51-66). New York: Springer.
- Thissen, D., Steinberg, L. & Fitzpatrick, A. R. (1989). Multiple choice model: The distractors are also part of the item. *Journal of Educational Measurement*, 26, 161-176.

Recebido em: fevereiro/2008

Revisado em: maio/2008

Aprovado em: junho/2008

Sobre os autores:

Facundo Juan Pablo Abal. Licenciado en Psicología. Ayudante de Trabajos Prácticos de Primera de la Cátedra II de Estadística de la Facultad de Psicología de la Universidad de Buenos Aires. Becario de Doctorado del Proyecto UBACyT P020. Investigador en el Proyecto PICT 20909 de la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT).

Gabriela Susana Lozzia. Licenciada y Profesora en Psicología. Jefe de Trabajos Prácticos Interina de la Cátedra II de Estadística de la Facultad de Psicología de la Universidad de Buenos Aires. Categoría Docente-Investigador V (cinco). Becaria de Doctorado de CONICET. Investigadora en los Proyectos P020 de la Universidad de Buenos Aires y PICT 20909 de la Agencia (ANPCyT).

María Silvia Galibert. Magister Scientiae en Biometría y Profesora de Matemáticas. Profesora Adjunta Regular de las Cátedra I y II de Estadística de la Facultad de Psicología de la Universidad de Buenos Aires. Categoría Docente-Investigador III. Codirectora del Proyecto P020 de la Universidad de Buenos Aires e Investigadora Integrante del Proyecto PICT 20909 de la Agencia (ANPCyT). Codirectora de Beca CONICET.

María Ester Aguerri. Magister Scientiae en Biometría y Licenciada en Ciencias Matemáticas. Profesora Adjunta Regular de las Cátedras I y II de Estadística de la Facultad de Psicología de la Universidad de Buenos Aires. Categoría Docente-Investigador III (tres). Codirectora del Proyecto P020 de la Universidad de Buenos Aires e Investigadora Integrante del Proyecto PICT 20909 de la Agencia (ANPCyT).

Horacio Félix Attorresi. Licenciado en Ciencias Matemáticas. Profesor Regular Titular de la Cátedra II de Estadística de la Facultad de Psicología de la Universidad de Buenos Aires. Categoría Docente-Investigador I (uno). Director de los Proyectos con Subsidio P020 de la Universidad de Buenos Aires y PICT 20909 de la Agencia (ANPCyT). Director de Becarios de UBACyT, CONICET y ANPCyT.