

APLICACIÓN DEL MODELO DE THISSEN & STEINBERG AL ANÁLISIS DE ÍTEMES DE UNA PRUEBA DE CRITERIO LÓGICO

APPLICATION OF THISSEN & STEINBERG'S MODEL TO THE ANALYSIS OF THE ITEMS OF A LOGICAL CRITERION TEST

Abal, Facundo¹; Lozzia, Gabriela²; Aguerri, M. Ester³; Galibert, M. Silvia⁴; Attorresi, Horacio⁵

RESUMEN

Se lleva a cabo el análisis de ítems de una prueba denominada de Criterio Lógico en el marco de la Teoría Clásica de Tests (TCT) y mediante el modelo politómico de Thissen y Steinberg de la Teoría de Respuesta al Ítem. Se comparan los resultados obtenidos y se confirma la hipótesis de que el modelo politómico es más exigente en cuanto a los requerimientos de calidad. Se procesaron los 30 ítems de esta prueba y los aceptables desde el punto de vista politómico se correspondían con los satisfactorios en la TCT mientras que algunos que podrían aceptarse desde la TCT evidenciaron fallas bajo el análisis politómico. Se ilustran con 3 ejemplos estas situaciones. Veintiséis ítems resultaron aceptables desde la TCT y sólo 7 cumplieron con los requerimientos de calidad del análisis politómico. El análisis de las curvas de las opciones permitió detectar aquéllas que podrían mejorarse, de qué manera y qué tipo de situaciones evitar en la construcción de los nuevos.

Palabras Clave: Modelo Politómico de Thissen y Steinberg, Análisis de Ítems, Teoría de Respuesta al ítem

ABSTRACT

Item analysis of a test named Logical Criterion was made within the frame of the Classical Test Theory (CTT) and by means of the Thissen & Steinberg's model of the Item Response Theory. The results obtained from both analysis are compared and the hypothesis of the polytomous model being more demanding as regards quality requirements is confirmed. The thirty items that make up this test were analyzed. As a result, those which were acceptable from the polytomous point of view were satisfactory also for the CTT, while some items that might be accepted in the CTT showed failures in the polytomous analysis. These situations are illustrated by three examples. Twenty six items turned out to be acceptable for the CTT but only seven of them achieved the requirements regarding quality that polytomous analysis needed. The analysis of the curves was useful to detect those that could be improved, as well as the way of improving them, and what kind of situations are to be avoided in the future construction of new items.

Key word: Thissen & Steinberg's Polytomous Model, Item Analysis, Item Response Theory

Instituto de Investigaciones. Fac. de Psicología, Universidad de Buenos Aires. La investigación que se presenta en este trabajo fue realizada con los subsidios de la Universidad de Buenos Aires (UBACyT P020 y P027) y del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET PIP Nro. 2426)

¹ Estudiante avanzado de la Lic. de Psicología. Ayudante de Segunda en Estadística, Cát II, Fac. de Psicología, UBA. Becario Estímulo del Proyecto UBACyT P020.

² Lic. y Prof. en Psicología, UBA. Ayudante de Trabajos Prácticos Regular en Estadística, Cát I y II, Fac. de Psicología, UBA. Investigadora del Proyecto UBACyT P020.

³ Prof. Adj. Reg. de Estadística, Cát: I y II, Fac. de Psicología, UBA. Codirectora de los Proyectos CONICET PIP 02426 y UBACyT BIENAL P020. M.Sc. en Biometría, UBA.

⁴ Prof. Adj. Reg. de Estadística, Cát: I y II, Fac. de Psicología, UBA. Codirectora del Proyecto UBACyT P020 e investigadora en el Proyecto CONICET PIP2426. M.Sc. en Biometría, UBA.

⁵ Prof. Reg. Titular de Estadística de la Fac. de Psicología, UBA. Director de los Proyectos UBACyT P020 y CONICET PIP2426

INTRODUCCIÓN

Como la buena calidad de un test se basa en la buena calidad de sus ítems, la etapa de análisis de ítems es crucial en el proceso construcción de una prueba. Bajo el enfoque de la Teoría Clásica de Test (TCT) los principales criterios a tener en cuenta en esta fase son: que la opción correcta (clave) haya sido más elegida que cualquier distractor (la moda de la distribución de frecuencias de las alternativas de cada ítem) y que la correlación ítem - test sea positiva. En cuanto a la distribución de frecuencias de los distractores, se prefiere que sea aproximadamente uniforme. Como las respuestas a los ítems suelen puntuarse de manera dicotómica con 1 ó 0 según sean correctas o incorrectas, la puntuación 0 no discrimina entre los distractores. Si entre ellos hubiera alguno más próximo a la opción correcta, elegirlo no contribuiría más al puntaje del sujeto que el haber elegido un distractor más lejano. La dicotomización, pues, pierde la información que podría estar contenida en los distractores. Esta pérdida no sería tal si todos ellos fueran "igualmente" lejanos porque no aportarían información diferente. La equidistancia se refleja en la homogeneidad de las frecuencias con las que son elegidos. Sin embargo, a la hora de inventar distractores, surgen naturalmente algunos más cercanos a la clave que otros y esto no sólo no debería ser un defecto del ítem sino que puede aportar información de interés desde el punto de vista cognitivo, si se discriminaran las diferentes respuestas erróneas. La Teoría de Respuesta a los Ítem (TRI) constituye una nueva perspectiva para la medición psicológica y educativa que ha hecho importantes contribuciones para enriquecer y complementar el análisis clásico de los ítems así como también a la comprensión de los procesos cognitivos involucrados en la respuesta al ítem. Sobre todo a través de los modelos de respuesta para ítems politómicos (van der Linden y Hambleton, 1997). Sus modelos parten de la premisa de que la respuesta de un sujeto a un ítem puede predecirse y explicarse a partir de una variable inobservable: el rasgo latente. A diferencia del enfoque clásico que se basa en resúmenes estadísticos globales, los modelos para ítems politómicos de la TRI permiten visualizar la respuesta de los sujetos a cada una de las opciones (describen la probabilidad de elección de cada alternativa) en función del nivel del rasgo que se desea medir. De este modo, no sólo es posible conocer cuántos eligieron cada alternativa sino también qué clase de sujetos, según su nivel de habilidad, se inclinaron por cada opción. Esta mayor riqueza de análisis del ítem puede aprovecharse no sólo para controlar la calidad psicométrica del instrumento sino inclusive para obtener información que podría ser de utilidad en un abordaje desde la perspectiva de un estudio cognitivo de los errores registrados. Otra aplicación importante es la orientación que brinda tanto para evitar errores en la construcción de nuevos ítems como para

reformular algunos de ellos.

La TRI permite descubrir que un ítem no es más confiable que otro sino para determinados niveles de habilidad pudiendo darse la situación contraria en otros niveles; todo ello depende de dónde alcancen los máximos sus correspondientes funciones de información.

La TRI ha alcanzado una gran difusión en EEUU y Europa pero en nuestro país es aún incipiente. Si bien existen algunos trabajos de aplicación de los modelos logísticos para ítems dicotómicos (Galibert et al., 1997, Aguerri et al., 1998, Cortada de Kohan, 1998, Attorresi et al., 1999) no se conocen publicaciones acerca de la aplicación de los modelos para respuestas politómicas en nuestro medio.

El objetivo de este trabajo es presentar una aplicación del Modelo "Multiple Choice" de Thissen y Steinberg al análisis de ítems de una prueba que Attorresi et al. (1993) denominaron de razonamiento de criterio lógico. Se exhiben algunos ejemplos que permiten comparar la utilidad de este modelo con respecto a la perspectiva clásica.

Modelos de Respuesta para Ítems de Elección Múltiple

Bock (1972) propuso un modelo para la respuesta $x = h$ con $h = 1, 2, \dots, m_j$ a un ítem j de elección múltiple de m_j alternativas:

$$P(x_j = h / \theta; \overset{\rho}{a}, \overset{\rho}{c}) = \frac{\exp(a_h \theta + c_h)}{\sum_{k=1}^{m_j} \exp(a_k \theta + c_k)} \quad \text{con} \quad \sum_{h=1}^{m_j} a_h = \sum_{h=1}^{m_j} c_h = 0$$

donde $\overset{\rho}{a}$ y $\overset{\rho}{c}$ son vectores m_j dimensionales sujetos a las restricciones con el fin de salvar la indeterminación de la escala y hacerlos estimables.

Samejima (1979) modificó el modelo de Bock (1972) incluyendo una categoría latente a la que denominó categoría cero; es la correspondiente a las respuestas de quienes eligen una opción por azar. En este modelo $d_h = 1/m_j$ indica la proporción de sujetos que, habiendo respondido por azar, eligen la alternativa h .

$$P(x_j = h / \theta; \overset{\rho}{a}, \overset{\rho}{c}) = \frac{\exp(a_h \theta + c_h)}{\sum_{k=0}^{m_j} \exp(a_k \theta + c_k)} + \frac{d_h \exp(a_0 \theta + c_0)}{\sum_{k=0}^{m_j} \exp(a_k \theta + c_k)}$$

donde $\overset{\rho}{a}$ y $\overset{\rho}{c}$ son vectores m_j+1 dimensionales sujetos a

$$\sum_{h=0}^{m_j} a_h = \sum_{h=0}^{m_j} c_h = 0$$

Thissen y Steinberg (1984, 1997) llaman DK: "don't know"

a la categoría 0 del modelo de Samejima y proponen una proporción d_h no constante para cada alternativa; por lo

que se agrega la restricción $\sum_{h=1}^{m_j} d_h = 1$. Así, la curva

DK representa la proporción de sujetos que contestan el ítem sin saber, en función de su nivel de habilidad y d_h la proporción de los mismos que eligió la alternativa h .

La respuesta "sin saber" puede pensarse como una respuesta "al azar" pero que no otorga, a diferencia del modelo de Samejima, la misma probabilidad a cada opción. Caracteriza a aquellos sujetos que, dada su falta de criterio o conocimiento, pueden ser atraídos más por una opción que otra pero por una razón no pertinente a la habilidad que mide el ítem; de allí que la proporción de elegir una u otra alternativa no es necesariamente homogénea.

El parámetro a se relaciona con el crecimiento y monotonía de las funciones. La alternativa a la que le corresponde el mayor valor de a suele ser la clave cuya curva se espera que sea monótona creciente y asintótica a 1 por la derecha. La curva DK comúnmente está asociada al valor más negativo de a , lo que implica, como es de desear, que sea decreciente y asintótica a cero por la derecha. Valores intermedios de a producen curvas no monótonas, que suelen ser las de los distractores. El parámetro c refleja la frecuencia relativa de elección de cada alternativa. Alternativas con similares valores de a serán más frecuentemente elegidas para mayores valores de c .

Características de la Prueba

Attorresi, H. y Pano, C. (1991) diseñaron una prueba de razonamiento lógico consistente de tres subpruebas cortas referidas a tres aspectos del mismo que dieron en llamar argumental, de criterio y de análisis. Attorresi, H., Pano, C.; Galibert, M. y Aguerri, M. (1993) describen el criterio lógico como un razonamiento universal, impreciso y de uso diario basado en el sentido común. Aparece como un sustento del discurso de la vida social y está más vinculado a lo que se entiende por filosofía práctica que a conocimientos técnicos complejos y estructurados. A los individuos con esta capacidad se los caracteriza como "criteriosos", "sensatos" o de "ideas claras". El criterio se desarrolla más por un ejercicio de reflexión interior o con charlas y discusiones cotidianas, antes que por un aprendizaje sistemático. Esto lo diferencia del razonamiento argumental al que se puede acceder a través del conocimiento de las reglas que rigen a la lógica o del analítico que se ejercita por la resolución de problemas donde se aplica principalmente la deducción.

La prueba de criterio lógico tiene como objetivo evaluar este tipo de razonamiento a través de la medición de

habilidades para:

- Relacionar correctamente azar, determinismo y aciertos, discriminando entre una decisión probable y racional y otra irracional o supersticiosa.
- Determinar las relaciones causales, causas generales y particulares entre variables de un cierto fenómeno económico, social, político o de otra índole descrito de forma elemental sin complejidad técnica.
- Distinguir, independientemente de las propias creencias, los objetos de la esfera de la ciencia y sus métodos, de los de la esfera de la fe y sus doctrinas.
- Analizar y construir estrategias para resolver o sacar conclusiones del relato simplificado de situaciones de la vida cotidiana o de un hecho policial
- Analizar y saber explicar ciertas conductas humanas inherentes a las relaciones familiares o a los conflictos de pareja.
- Diferenciar posiciones éticas de criterios pragmáticos.
- Precisar el significado de una polisemia en virtud de su mejor ajuste al contexto y discriminar entre utilidad, preferencia y necesidad.
- Interpretar con objetividad un mensaje evitando interferencias de los propios prejuicios o posición ideológica.

Esta enunciación de habilidades no es exhaustiva pero caracteriza el universo de aplicaciones de este razonamiento, el cual provee contenidos para confeccionar los ítemes.

Para resolver un ítem de criterio lógico es necesario: 1.- Comprender conceptual, lógica y semánticamente el texto con que se formula el ítem. 2.- Conocer el significado preciso y diferencial de los términos que se usan. 3.- Tomar muy en cuenta el contexto y la consigna propia de cada ítem, como por ejemplo no es lo mismo "Señale la opción correcta" que "Usted diría que...". 4.- Dejar de lado las opiniones, creencias, simpatías o antipatías propias para sólo decidir en función de lo que plantea la base del ítem y de su comparación con las opciones dadas. 5.- Saber desechar las opciones que contengan elementos contradictorios con la base del ítem, como así también la que no muestra relación con ella. 6.- Encontrar algún criterio que se apoye en la lógica del discurso presentado y que pueda discriminar entre las opciones.

METODOLOGÍA

Muestra. La muestra está integrada por 998 estudiantes de segundo año de la carrera de Psicología de la Universidad de Buenos Aires distribuidos en todas las bandas horarias. El 85.3% fueron mujeres y la edad media fue de 23 años (SD = 6.29) La mediana de edad fue de 20 años con un mínimo de 18 y un máximo de 64 años.

Instrumento. El instrumento consta de 30 ítemes de cuatro opciones de respuesta. Ocho de ellos corresponden

a la prueba original aunque fueron modificados en función de las recomendaciones de los jueces y resultados obtenidos en pruebas pilotos. Para disminuir el efecto de la fatiga se construyeron dos cuestionarios con ordenamiento inverso.

Cada ítem consiste en un encabezado basado en algún dicho, opinión o explicación. A partir de éste, se debe tratar de encontrar, entre las opciones, la conclusión pedida o la afirmación que la sustenta. Se reconoce la posibilidad de que todas las respuestas pueden ser correctas, pero sólo una de ellas es la mejor. En el instructivo se sugiere como estrategia de resolución que el sujeto comience por descartar las opciones que menos se ajusten a lo pedido y finalmente discrimine entre las restantes la que responde con mayor exactitud a la consigna.

Procedimiento. Se realizó el análisis de ítemes tradicional en el marco de la TCT. Se obtuvieron las distribuciones de frecuencias para cada opción con el programa STATISTIX™ y se obtuvieron indicadores de confiabilidad (alpha de Cronbach) y la correlación ítem-test con TESTAT de SYSTAT™.

Se utilizó el programa MULTILOG™ (Thissen y Steinberg, 1991) para estimar los parámetros de las curvas a partir del modelo de Thissen y Steinberg. Luego, a partir de estos parámetros, se obtuvieron las curvas de respuesta correspondientes a cada una de las opciones de los treinta ítemes de criterio lógico. También se obtuvo la curva "Don't Know". Las curvas se graficaron operando con el programa STATISTIX™. A partir de ellas, se llevó a cabo el análisis de ítemes.

RESULTADOS

Desde la perspectiva clásica se puede apreciar que todas las correlaciones ítem-test fueron positivas y oscilaron entre 0.143 y 0.368. Se obtuvo un coeficiente alpha de 0.438. El 87% de las distribuciones de frecuencias de los ítemes tuvo a la respuesta correcta como moda. Los índices de facilidad fueron desde 19.3 hasta 73.3.

Del análisis de las curvas obtenidas del modelo de Thissen y Steinberg se puede observar que: en 16 de los 30 ítemes la curva de la respuesta correcta es creciente, de los cuales siete respondieron al comportamiento deseable de tener las curvas de todos los distractores decrecientes y el resto tiene sólo una curva de distractor creciente.

En los ítemes fáciles, sería conveniente modificar las alternativas para que sus curvas no resulten nulas y aporten más información.

A continuación se ejemplificarán diversas situaciones que se presentaron en el análisis de tres ítemes. Se hará referencia a las figuras para visualizar sus curvas y a las tablas para los resúmenes estadísticos.

Ejemplo de ítem con comportamiento regular en el análisis de ítemes de la TCT y curvas no satisfactorias en el análisis politómico

Ítem 8) Dice Federico: "Mis amigos Alberto y Gustavo jugaron al Loto en la agencia 'Las Vegas' y ganaron entonces..."

¿A qué conclusión razonable arribaría ud.?

- a) Se ha dado una feliz coincidencia.
- b) A Federico le conviene jugar en la agencia "Las Vegas".
- c) Generalmente los que ganan al Loto han jugado en la agencia "Las Vegas".
- d) A Federico le conviene que Alberto y Gustavo elijan los números.

Para resolver este ítem se requiere la habilidad para discriminar una conducta o decisión racional de una irracional o supersticiosa.

Si bien la correlación ítem-test es relativamente baja ($r = 0.187$) y su eliminación haría aumentar el coeficiente alpha a 0.441, la distribución de frecuencias es aceptable. La Tabla 1 nos muestra que la opción a) (que es la correcta) ha resultado ser la más atractiva ya que fue elegida por el 59% de los sujetos.

A pesar de haber obtenido resultados regulares desde la TCT, el análisis politómico puede ampliar la descripción de este ítem y encontrar sus fallas. La Figura 1 permite apreciar que la curva de la respuesta correcta (CC) aumenta ligeramente de 0.5 a 0.6 desde los niveles más bajos de habilidad hasta el nivel medio, pero retorna al primer valor en los niveles más altos. Esto quiere decir que dentro del porcentaje de sujetos que han contestado correctamente, no se ha discriminado ningún nivel de habilidad. Como se observa en la Tabla 1, el ítem no otorga información relevante en ningún nivel de habilidad.

La curva DK ha resultado prácticamente nula. Esto implica que es bastante probable que todos los sujetos, sin importar su nivel de habilidad, realizaron su elección deliberadamente.

Las opciones c) y d) poseen curvas decrecientes siendo el último el distractor más elegido en los niveles de habilidad más bajos. Resulta llamativo el crecimiento de la curva del distractor b) el cual no tiene nada de razonable. Ésta nace con una probabilidad inferior a 0.1 y hacia los niveles más altos de habilidad alcanza una probabilidad superior a 0.5; superando, inclusive, a la respuesta correcta.

Si bien la TCT nos informa que es un ítem de regular calidad, a partir del análisis de los ítemes se evidencia en qué puntos se encuentran las fallas.

Ejemplo de ítem con un buen comportamiento en el análisis de ítemes de la TCT y curvas no satisfactorias

en el análisis politómico

Ítem 20) Lucas es perezoso y adeuda varias materias de 1er. año de la escuela. A pesar del riesgo de repetir y de los esfuerzos de sus padres para persuadirlo de que se ponga a estudiar, se deja estar arguyendo que le alcanzará el tiempo.
 Su madre le dice: *“Lamentablemente recién vas a reconocer que yo tenía razón cuando rindas mal las materias. Entonces será demasiado tarde porque habrás repetido el año”*.
 Al hablar de esa manera la madre busca lograr el siguiente efecto:

- a) que si Lucas repite el año reconozca que ella tenía razón y comience a obedecerla.
- b) que Lucas adquiera la madurez suficiente para salir adelante sin que le anden atrás.
- c) que Lucas no tenga que atravesar por una experiencia frustrante para comenzar a madurar.
- d) que Lucas se ponga a estudiar inmediatamente para que no pierda el año.

Para resolver este ítem se requiere la habilidad para analizar y saber explicar ciertas conductas humanas inherentes a las relaciones familiares.

Este ítem ha demostrado excelentes propiedades desde la TCT: es muy satisfactoria la distribución de frecuencias entre las opciones (Tabla 2), su eliminación haría disminuir el coeficiente alpha a 0.407 y la correlación ítem-test es 0.368, el índice más elevado.

En este ítem la CC se mantuvo aproximadamente constante en 0.4 en los niveles bajos de habilidad, a partir del cual comenzó a crecer lentamente hasta alcanzar una probabilidad cercana a 0.6. Nótese además en la Figura 2 que la curva DK es muy alta en los niveles bajos cayendo abruptamente luego. A partir de $\theta = -0.7$ la CC comienza a ser más probablemente elegida y decae abruptamente la probabilidad de elección sin saber. La Tabla 2 muestra un índice de respuesta por azar de 0.17 correspondiéndole un 0.41 a la opción correcta. De ambas fuentes podemos sintetizar que el 17% de los sujetos contestaron al azar, los cuales corresponden casi exclusivamente a los niveles bajos de habilidad. De ellos, el 41% eligió la opción correcta. En otras palabras, en este caso el elegir la opción correcta no da cuenta de un buen nivel de habilidad. Por esa razón este ítem debe reformularse.

La CC es relativamente “horizontal” lo que indica que este ítem presenta una baja discriminación para todos los niveles de habilidad de los sujetos.

Es de notar la tendencia constante de la opción c), el distractor más elegido en los niveles de habilidad medios y altos. Tal vez en el proceso de reflexión, los sujetos hayan sido más propensos a seleccionar esta opción por ser estudiantes de psicología ya que hace referencia a

términos psicológicos, dando una explicación “psicologista” de la situación. Nótese que esta característica no se hubiera detectado con el análisis clásico.

Ejemplo de ítem con un buen comportamiento en el análisis de ítems de la TCT y curvas satisfactorias en el análisis politómico

Ítem 6) Un cliente le dice al vendedor: *“Yo compraría el sillón bordó, aunque en cuestiones de decoración, mi esposa tiene la última palabra”*.
 Señale la opción correcta.

- a) El cliente deja supeditada su decisión a la opinión de su esposa.
- b) El cliente no confía en su gusto para decorar su casa.
- c) El cliente espera escuchar la opinión de su esposa.
- d) El cliente expresa su opinión independientemente de lo que opine su mujer.

Para resolver este ítem se requiere la capacidad de extraer conclusiones de un relato simplificado de situaciones de la vida cotidiana.

Los índices clásicos resultan satisfactorios: la correlación ítem-test es 0.365, su extracción haría disminuir el coeficiente alpha a 0.407 y el porcentaje de respuesta correcta (60%) es superior al de cualquiera de los distractores, para los cuales hay una buena distribución de las frecuencias.

La probabilidad de seleccionar la respuesta correcta fue siempre mayor que la correspondiente a las de los distractores. La CC es creciente desde los niveles medios de habilidad y asintótica a 1 por la derecha, lo cual responde al comportamiento deseado.

La Figura 3 muestra que los sujetos con niveles de habilidad por debajo de la media son atraídos por todas las alternativas de respuesta mientras que los más hábiles por la opción correcta a). Las opciones b) y d) aunque son decrecientes son de baja discriminación entre los diversos niveles de habilidad.

La curva de la opción c) (el distractor más cercano) se mantuvo aproximadamente constante en 0.2 desde los niveles bajos hasta el nivel medio de habilidad, a partir del cual comenzó a ser decreciente.

La curva DK fue estrictamente decreciente a lo largo de todo el rango de habilidad, siendo relativamente alta desde los niveles bajos hasta los intermedios (0.8 a 0.3 respectivamente). La curva de respuesta al azar ha favorecido a los sujetos con bajo nivel del rasgo como lo indica la fila DK de la Tabla 3; 62% ha elegido la opción correcta sin saber.

Tanto el análisis politómico (Figura 3) como el clásico (Tabla 3) muestran que es un ítem de dificultad intermedia. Su mayor información la brinda en el nivel de habilidad

0,5 por lo que claramente discrimina mejor a los sujetos de niveles medios a altos de habilidad. Es un ejemplo de ítem con buen funcionamiento tanto en términos de los índices clásicos como del análisis politómico.

CONCLUSIONES

El modelo Thissen y Steinberg muestra las limitaciones que surgen de un análisis global de los ítems a partir de la TCT. Éste último resulta útil, pero no exhaustivo si se pretende conocer en profundidad la calidad de los mismos. Por ello, gran cantidad de ítems de esta prueba han resultado aceptables desde la TCT y sólo siete han cumplido con los requerimientos de calidad del análisis politómico.

El análisis de las curvas de las opciones para cada ítem permitió detectar aquéllas que podrían mejorarse, qué orientación seguir en el mejoramiento del ítem y qué tipo de situaciones evitar en la construcción de los nuevos. También se observa que no se han conseguido ítems claramente discriminatorios, en particular para los niveles medios de habilidad.

Obviamente en algunos casos no puede pretenderse un comportamiento óptimo de los ítems respecto de todas las variables consideradas, so riesgo de descartar demasiado material. Por ejemplo, difícilmente todos los distractores de un ítem fácil representarán una oposición real a la correcta; muy probablemente alguno capte muy poco la atención de los sujetos. Como esto sucedió en la mayoría de los ítems de esta prueba, se propone eliminar un distractor para una próxima versión. Tampoco siempre se puede lograr distractores que discriminen en todos los niveles de habilidad; siendo una situación frecuente que algún distractor es elegido con igual probabilidad por los sujetos de todas las clases. La aceptación de un ítem supondrá pues, en algunos casos, resignar algunas características deseables en tanto se mantengan las propiedades fundamentales.

Aunque el análisis politómico persigue como finalidad principal la buena calidad de los ítems, eventualmente también puede suscitar la formulación de hipótesis de interés psicológico. En efecto, al intentar dar razones que expliquen qué pudo haber movido a los sujetos de tales o cuales niveles de habilidad a elegir más a este distractor que a aquél, pueden generarse hipótesis o plantearse problemas interesantes para ser abordados desde la perspectiva de un estudio cognitivo.

BIBLIOGRAFÍA

Aguerri, Ma Ester; Galibert, Ma Silvia; Veneri, Mónica y Attorresi, Horacio: Teoría Clásica de Test y Teoría de Respuesta al Ítem: un Estudio Comparativo aplicado a una Prueba de Razonamiento Verbal. VI Anuario de Investigaciones 1998. Facultad de

- Psicología, UBA. Buenos Aires, Argentina. pp 68-78.
- Attorresi, H. y Pano, C. (1991): Aportes para el diseño de una prueba de razonamiento lógico para egresantes del ciclo medio de enseñanza. Anuario de Investigaciones. Volumen 2, año 1990/91. Facultad de Psicología, UBA. Buenos Aires, Argentina. pp. 11-14.
- Attorresi, H.; Pano, C.; Galibert, M. y Aguerri, M. (1993): Diseño, construcción y análisis de una prueba de criterio lógico. Anuario de Investigaciones. Volumen 3, año 1992/93. Facultad de Psicología, UBA. Buenos Aires, Argentina. pp. 19-26.
- Attorresi, Horacio; Aguerri, M. Ester y Galibert, M. Silvia (1999): Aplicación del modelo logístico de tres parámetros en una prueba de completar frases. Investigaciones en Psicología, Año 4 Nro1. 1999. Buenos Aires, Argentina. pp 7-25
- Bock, R. (1972). Estimating item parameters and latent ability when responses are scored in two or more latent categories. *Psychometrika* 37, 29-51.
- Cortada de Kohan, Nuria (1998). "La Teoría de Respuesta al Ítem y su aplicación al 'Test Verbal Buenos Aires'". *Interdisciplinaria, Revista de Psicología y Ciencias Afines*. Vol. 15, nros.1-2, 101-129.
- Galibert, M.; Aguerri, M.; Attorresi, H.; Pano, C.; Nicolai, L. y Kiel, L. (1997). La Teoría de Respuesta al Ítem y su aplicación a una prueba de razonamiento verbal. *Vº Anuario de Investigaciones*, 365-379. Facultad de Psicología, UBA.
- Martínez Arias, R. (1995). *Psicometría: Teoría de los Tests Psicológicos y Educativos*. Madrid. Síntesis.
- Muñiz Fernández, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid. Ediciones Pirámide.
- Muñiz, J. (1996b). *Psicometría*. Madrid, España. Editorial Universitas, S.A.
- 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. y Steinberg, L. (1991). *MULTILOG™ User's Guide. Multiple, Categorical Item Lanalysis and Test Scoring Using Item Response Theory*. Scientific Software, Inc.
- Thissen, D. y 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*. NewYork, Springer.
- Thissen, D., Steinberg, L., y Fitzpatrick, A. (1989). Multiple choice models: The distractors are also part of the item. *Journal of Educational Measurement* 26, 161-176.
- Van der Linden, W. & Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York, Springer.

Tabla 1: Fragmento de salida traducida de MULTILOG

Ítem 8 Clave: a

Estimación de los parámetros para la obtención de curvas de respuesta									
Categoría(K)	DK	a	b	c	d				
A(K)	-3.36	0.94	1.50	0.61	0.32				
C(K)	-8.22	3.35	2.33	1.04	1.50				
D(K)		0.02	0.28	0.69	0.01				
Máximo de información que aporta el ítem para estimar cada nivel de habilidad									
THETA	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
I(THETA):	0.13	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.10
Distribución de frecuencias de las opciones									
Categoría(K)	DK	a	b	c	d				
Frec. Obs.	0	579	233	64	110				
Prop. Obs.	0.00	0.59	0.24	0.06	0.11				
Prop. Esp.	0.00	0.59	0.24	0.06	0.11				

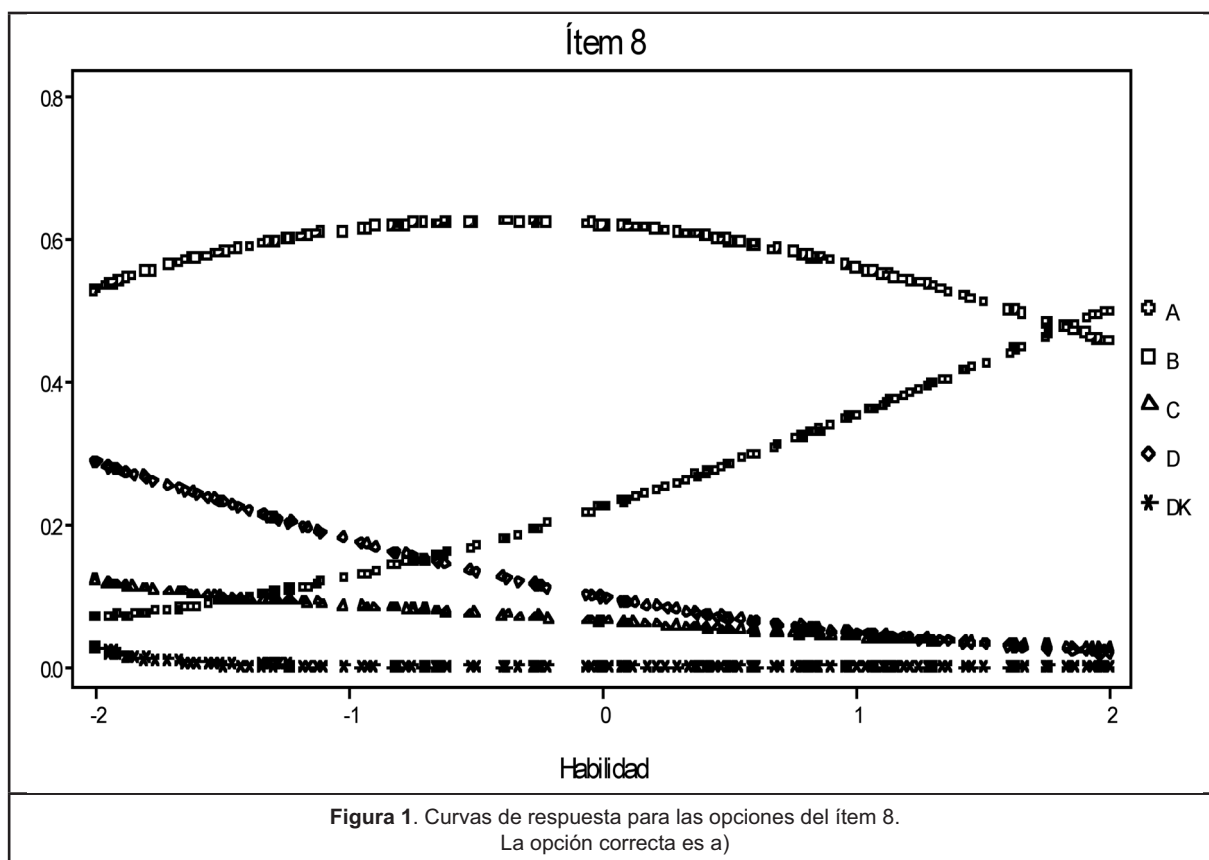


Tabla 2: Fragmento de salida traducida de MULTILOG
 Ítem 20 Clave: d

Estimación de los parámetros para la obtención de curvas de respuesta					
Categoría(K)	DK	a	b	c	d
A(K)	-5.79	1.63	0.70	1.70	1.75
C(K)	-3.77	0.39	0.08	1.39	1.90
D(K)		0.29	0.07	0.24	0.41

Máximo de información que aporta el ítem para estimar cada nivel de habilidad									
THETA	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
I(THETA):	0.00	0.00	0.49	0.13	0.08	0.05	0.03	0.02	0.01

Distribución de frecuencias de las opciones					
Categoría(K)	DK	a	b	c	d
Frec. Obs.	0	139	73	287	486
Prop. Obs.	0.00	0.14	0.07	0.29	0.49
Prop. Esp.	0.17	0.14	0.07	0.29	0.49

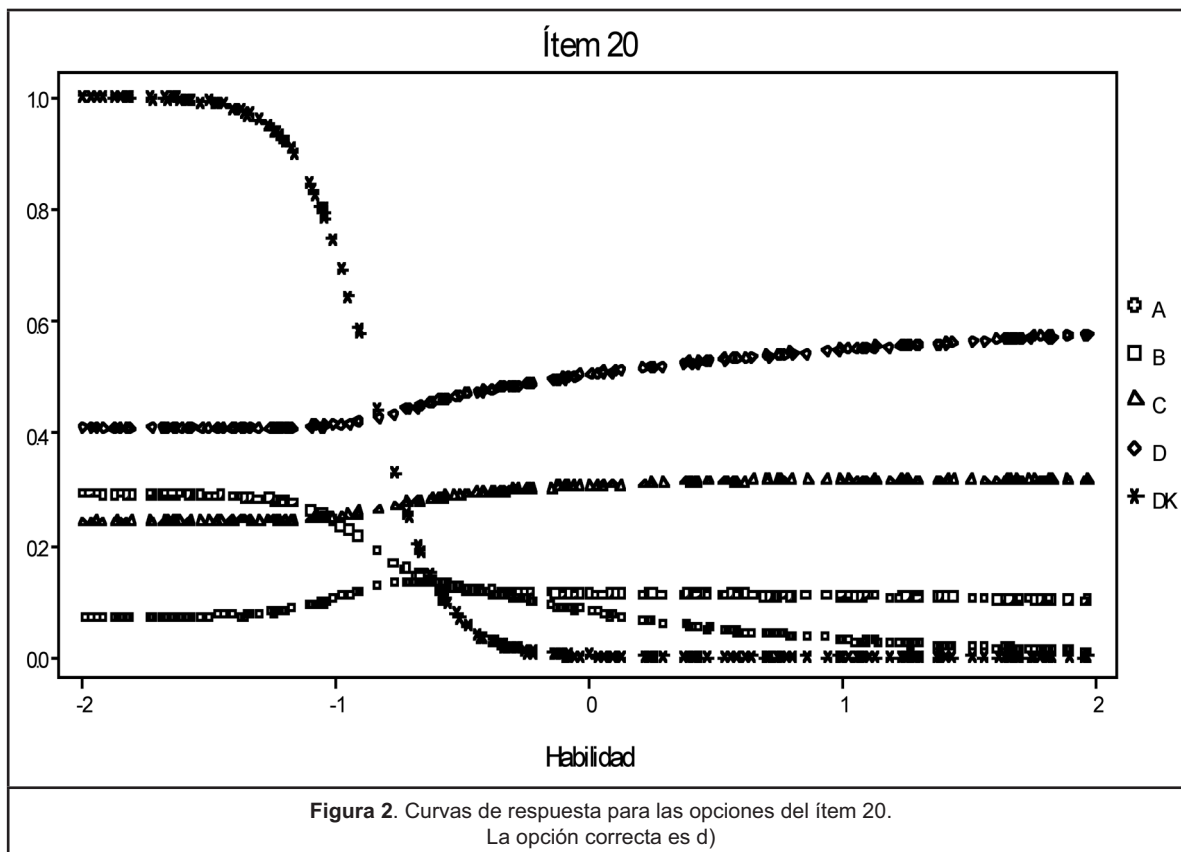


Tabla 3: Fragmento de salida traducida de MULTILOG
Ítem 6 Clave: a

Estimación de los parámetros para la obtención de curvas de respuesta									
Categoría(K)	DK	a	b	c	d				
A(K)	-0.63	1.58	-0.76	0.41	-0.60				
C(K)	0.93	0.62	-0.69	-0.53	-0.34				
D(K)		0.62	0.01	0.35	0.02				
Máximo de información que aporta el ítem para estimar cada nivel de habilidad									
THETA	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
I(THETA):	0.00	0.01	0.02	0.08	0.27	0.45	0.37	0.20	0.09
Distribución de frecuencias de las opciones									
Categoría(K)	DK	a	b	c	d				
Frec. Obs.	0	593	82	202	110				
Prop. Obs.	0.00	0.60	0.08	0.20	0.11				
Prop. Esp.	0.37	0.60	0.08	0.20	0.11				

