

UNA PROPUESTA PARA LA OBTENCIÓN DE NIVELES DE DESEMPEÑO EN LOS MODELOS DE TEORÍA DE RESPUESTA AL ÍTEM*

Víctor H. Cervantes** & Edilberto Cepeda-Cuervo***
Universidad Nacional de Colombia, Colombia

Sandra Liliana Camargo****
Universidad de los Andes, Colombia

Resumen

En este artículo, luego de una breve presentación de los modelos de la Teoría de Respuesta al Ítem, se introduce el proceso de anclaje de la escala presentado por Beaton y Allen (1992) para la interpretación de escalas y se propone un procedimiento para obtener los niveles de desempeño necesarios para el mismo. El procedimiento propuesto es empleado en la interpretación de una escala del cuestionario ‘Personas en mi vida’ (Cook, Greenberg & Kusche, 1995; Camargo, Mejía, Herrera & Carrillo, 2007).

Palabras clave: Teoría de Respuesta al Ítem, anclaje de la escala, niveles de desempeño, TRI, apego.

Abstract

In this article, after a brief presentation of the Item Response Theory, it is introduced the scale anchoring process presented by Beaton and Allen’s (1992) for scales interpretation; and a procedure to obtain the needed performance levels is introduced. The proposed methodology is used to interpret a “People in my life” questionnaire scale (Cook, Greenberg & Kusche, 1995; Camargo, Mejía, Herrera & Carrillo, 2007).

Key words: Item Response Theory, scale anchoring, performance levels, IRT, attachment.

Introducción

En diversos campos, en especial en las ciencias sociales, es frecuente que el interés dentro de una investigación recaiga sobre el estudio de características que no son observables directamente, sino inferidas a partir de otras que sí lo son. Por ejemplo, hay características como el nivel de calidad de vida de una persona que padece cáncer, o cómo es la relación afectiva entre un niño y sus padres, que al no poder ser medidas directamente requieren la adquisición de información obtenida de otras variables, como el nivel de dolor percibido. En estos casos, la información que se puede obtener es vista como un indicador de la característica –v.g. sobre un niño que siente confianza en la exploración de un sitio con el que no está familiarizado y busca a su madre con periodicidad, se puede decir que demuestra seguridad en su relación de apego con ella. En general, esta información se recolecta empleando instrumentos de observación de las conductas de las personas o cuestionarios en que se pregunta por aspectos relacionados con la variable de interés. Es claro que para estas variables, al no ser observables directamente, no se posee una escala definida e interpretable de antemano (Bartholomew, 1980, 1984; Ferrando, 1996).

* Este trabajo fue realizado por el primer autor durante su disertación de maestría bajo la dirección del segundo y con la colaboración del tercero. Los autores agradecen a Camargo, Mejía, Herrera y Carrillo (2007) quienes facilitaron las bases de datos para la realización de este trabajo.

** Departamento de Estadística, Universidad Nacional de Colombia, Bogotá-Colombia. E-mail: vhcervantesb@unal.edu.co

*** Departamento de Estadística, Universidad Nacional de Colombia, Bogotá-Colombia. E-mail: ecepedac@unal.edu.co

**** Departamento de Psicología, Universidad de los Andes, Bogotá-Colombia. E-mail: sa-camar@uniandes.edu.co

Con el fin de analizar y modelar estas variables no observables a partir de variables observadas, aparecen los modelos de rasgo o estructura latente (Lord, 1950; Lazarsfeld, 1955; Birnbaum, 1968), también conocidos como modelos de respuesta al ítem (Muñiz, 1996, 1997). En estos modelos se busca representar con claridad la relación entre los diversos reactivos, ítems o conductas observadas con la variable no observable de interés, también llamada variable latente. Concretamente, se establece una relación funcional entre la probabilidad de que un sujeto dé cierta respuesta a un reactivo y la magnitud que él o ella posee en la variable latente (Lord, 1980).

La interpretación de la escala que proveen los modelos de respuesta al ítem es uno de los principales asuntos prácticos que deben abordar el constructor de instrumentos, el profesional que aplica e interpreta cuestionarios de evaluación y la persona que responde a los mismos. El primer paso para lograr esta interpretación es la estimación de los parámetros involucrados en el modelo (Lord, 1986). El segundo paso es dar cuenta de qué representan los diferentes valores de la escala de medición y que posteriormente serán asignados a un sujeto a quien se evalúan sus capacidades o preferencias mediante la aplicación del instrumento. Una herramienta interesante que ha adquirido importancia en la interpretación de evaluaciones nacionales de competencia es el procedimiento de anclaje de la escala (Beaton & Allen, 1992; Beaton & Johnson, 1992; Andrade, Tavares & Valle, 2000; Pinto, 2006). Al emplear un instrumento específico, p.e. una prueba psicológica, no es posible dar cuenta de todos los posibles valores de la escala, por lo cual en el procedimiento propuesto por Beaton y Allen (1992) se procede a la selección a priori de algunos valores a los cuales se intenta asignar un significado respecto al constructo evaluado. Sin embargo, la selección apropiada de qué valores de la escala tomar, o incluso, la exploración de cuáles son los valores a los cuales el instrumento en verdad permite asignar una interpretación sustantiva respecto al constructo, es un problema que no ha sido abordado en la literatura.

En este trabajo se presenta una propuesta para la determinación empírica de los niveles de desempeño requeridos para el procedimiento de anclaje de la escala, con el fin de aportar a la solución de este problema en el contexto de pruebas con ítems dicótomos. En la primera sección se hace una breve presentación de la Teoría de Respuesta al Ítem en la cual se enmarca la propuesta; en la segunda, se plantea el problema de la interpretación de la escala de habilidad, se presentan el procedimiento de anclaje de la escala introducido por Beaton y Allen (1992) y una propuesta para la obtención de los niveles de desempeño; en la tercera, se ilustra la utilización del anclaje de la escala con una aplicación de la propuesta al análisis de una escala del cuestionario '*Personas en mi vida*'; finalmente, se discuten las ventajas y dificultades de la utilización del procedimiento planteado.

Aspectos generales de los modelos de Teoría de Respuesta al Ítem

El supuesto básico de los modelos de la Teoría de Respuesta al Ítem (TRI) es que existe una relación funcional entre la magnitud en la característica latente y la probabilidad de que un individuo dé o presente cierta respuesta en las variables observadas. En estos modelos se acostumbra denominar atributo o habilidad a la variable latente de interés, y reactivo o ítem a cada una de las variables observadas (Hambleton, Swaminathan & Rogers, 1991; Andrade et al., 2000; Baker & Kim, 2004); esta convención se seguirá en el resto de este documento. El segundo supuesto de los modelos TRI es la independencia local entre los ítems; es decir que para un individuo con cierto nivel de habilidad, la respuesta que da a un ítem es estadísticamente independiente de la respuesta que da a otro ítem (Lord & Novick, 1968; Hambleton et al., 1991; Muñiz, 1997; Andrade et al., 2000).

De acuerdo con qué supuestos adicionales se impongan sobre la magnitud del atributo (θ), las posibles respuestas a los ítems (U) y cómo se relacionan estas dos (f), se obtienen diferentes modelos TRI. Los modelos más utilizados consideran que θ es un número entre $-\infty$ e ∞ , y U puede tomar valores 0 o 1 (0 incorrecto y 1 correcto, respectivamente), es decir, son modelos en los que el atributo es unidimensional y los ítems son dicótomos (Andrade et al., 2000). En estos casos, la gráfica de la función f , es

frecuentemente llamada Curva Característica del Ítem (CCI) (Hambleton et al., 1991; Linden & Hambleton, 1997). Finalmente, de acuerdo con la forma y parametrización específica de la función f , surgen diferentes modelos. Los modelos más utilizados en la evaluación de atributos unidimensionales, para los cuales se han observado reactivos dicótomos, son los modelos logísticos; los principales son los de uno, dos y tres parámetros, aunque el más general es el modelo de tres parámetros (3-PL). Los otros se pueden obtener como casos especiales del modelo 3-PL. El modelo 3-PL (Birnbaum, 1968; Mislevy & Bock, 1997; Baker & Kim, 2004) está dado por:

$$P(U_{ij} = 1 | \theta_j, \alpha_i, b_i, c_i) = c_i + \frac{(1 - c_i)}{1 + e^{-D\alpha_i(\theta_j - b_i)}} \quad (1)$$

con $i = 1, 2, \dots, I$ y $j = 1, 2, \dots, n$, donde U_{ij} es una variable dicótoma que toma el valor 1 cuando el individuo j responde correctamente al ítem i , y 0 si no responde en forma correcta; θ_j y b_i se encuentran en la misma escala y toman valores entre $-\infty$ e ∞ , aunque usualmente toman valores entre -4 y 4 , además, en el punto b_i se tiene el punto de inflexión en la CCI; α_i es proporcional a la pendiente de la CCI en el punto b_i ; c_i indica la probabilidad de que un individuo responda correctamente al ítem i cuando su valor de atributo tiende a menos infinito, y D es un factor de escala constante para todos los ítems y toma valores 1 ó 1.7. De estas propiedades de los parámetros se derivan las siguientes interpretaciones para cada uno de ellos: θ_j representa la magnitud de atributo del individuo j ; α_i es el parámetro llamado de discriminación del ítem i ; b_i corresponde al parámetro de dificultad, también llamado de umbral, del ítem i ; c_i es conocido como el parámetro de adivinación o pseudoazar. En cuanto a D , cuando toma el valor 1.702, los demás parámetros son prácticamente idénticos a los correspondientes de los modelos de ojiva normal (Birnbaum, 1968).

El modelo logístico de dos parámetros se obtiene cuando se toma c_i igual a cero; el modelo de un parámetro, cuando además se toma α_i igual a uno, en los I ítems considerados. En el modelo 3-PL no se considera propiamente la función de probabilidad logística, pero en los modelos de uno y dos parámetros sí; por esta razón el nombre también es utilizado en el modelo de tres parámetros (Linden & Hambleton, 1997).

Anclaje de la escala para la interpretación de las habilidades

Interpretación de la escala de habilidades

Como se anotó previamente, la escala de la variable latente no está definida de antemano; por consiguiente, no es interpretable de forma natural. Una vez estimados los parámetros, es necesario dar una interpretación a la escala de habilidades obtenida. Este problema se aborda dando un sentido a las características de los individuos con cierto nivel de atributo, tomando como referencia aquel conjunto de ítems que mejor describe dicho nivel (Wright & Stone, 1979/1998; Andrade et al., 2000).

En Wright y Stone (1979/1998) se encuentra un método que busca describir el continuo de la escala latente para el modelo de Rasch, basado en los errores estándar de las estimaciones de los parámetros b_i . Beaton y Allen (1992) presentan un método de anclaje de la escala que procura describir diferentes puntos en la escala de habilidades teniendo en cuenta la capacidad de los ítems para discriminar entre dichos puntos. En este texto se manejará el procedimiento de anclaje de la escala como base para dar interpretación a la escala de habilidades (Beaton & Allen, 1992; Andrade et al., 2000).

Procedimiento de anclaje de la escala

El anclaje de la escala requiere la selección de varios niveles de desempeño sobre la métrica de las estimaciones de los parámetros de los ítems, o sobre alguna transformación lineal de los mismos (Beaton &

Allen, 1992; Beaton & Johnson, 1992; Andrade et al., 2000). El procedimiento consiste en seleccionar los ítems en la vecindad de cada nivel de desempeño que mejor diferencian entre los individuos en ese nivel y el nivel previo. Para esto, proponen los siguientes criterios sobre las probabilidades de respuesta correcta de cada nivel de habilidad θ_p :

$$\bullet P(U = 1 | \theta_p) \geq 0.65 \quad (2)$$

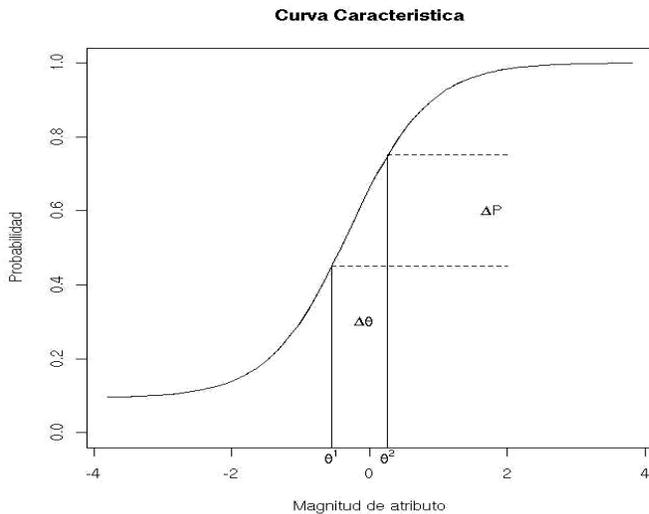
$$\bullet P(U = 1 | \theta_{p-1}) < 0.5 \quad (3)$$

$$\bullet P(U = 1 | \theta_p) - P(U = 1 | \theta_{p-1}) \geq 0.3 \quad (4)$$

Se dice que un ítem U pertenece al nivel de desempeño θ_p si y sólo si cumple los tres criterios anteriores (Beaton & Allen, 1992; Andrade et al., 2000; Pinto, 2006). Con estos criterios cada ítem sólo puede pertenecer a uno o a ningún nivel de desempeño, y puede presentarse el caso de que ningún ítem resulte anclado a los niveles de desempeño propuestos (Beaton & Allen, 1992). El principal inconveniente que presenta este procedimiento proviene de la necesidad de fijar los niveles de desempeño con anticipación.

Propuesta para la selección empírica de niveles de desempeño

A continuación se presenta un procedimiento empírico para la selección de los niveles de desempeño θ_p , $p = 0, 1, \dots, P$, para el conjunto de ítems. Este procedimiento se inspira en el presentado por Pinto (2006) y en el procedimiento de construcción de una variable, presentado por Wright y Stone (1979/1998,



$$\Delta\theta = \theta_2 - \theta_1, \quad \Delta P = P(U_i = 1 | \theta_2) - P(U_i = 1 | \theta_1)$$

Figura 1. Criterios de anclaje.

Cap. 4). Los criterios propuestos por Beaton y Allen (1992) para anclar un ítem de la escala se ilustran en la Figura 1 a partir de la pareja ordenada de puntos (θ_1, θ_2) , donde la primera condición concierne a $P(U = 1 | \theta_2)$, la segunda a $P(U = 1 | \theta_1)$ y la tercera a ΔP , la diferencia entre las dos probabilidades anteriores. Se puede apreciar que para un ítem es posible buscar un par de puntos en la escala de habilidad, llamémoslos θ_i y θ_h , que cumplan los criterios necesarios para el anclaje y tales que cualquier pareja ordenada de puntos (θ_l, θ_r) en la que $\theta_l \leq \theta_i$ y $\theta_r \geq \theta_h$ permitiría anclar al ítem en el nivel θ_2 respecto al nivel θ_1 , y en la que $\theta_l > \theta_i$ y $\theta_r < \theta_h$ no permitiría anclar al ítem. Además, la pareja (θ_i, θ_h) proporciona el intervalo más corto en la escala de habilidades para el cual es posible anclar al ítem.

Es fácil ver que en los modelos de uno y dos parámetros, la pareja (θ_i, θ_h) está dada por los valores de θ tales que $P(U_i = 1 | \theta_i) = 0.35$ y $P(U_i = 1 | \theta_h) = 0.65$, respectivamente. En el modelo de tres parámetros este intervalo está dado por los valores de θ tales que:

$$P(U_i = 1 | \theta_i) = \frac{c_i + 1}{2} - 0.15 \text{ y } P(U_i = 1 | \theta_h) = \frac{c_i + 1}{2} + 0.15, \text{ si } c_i < 0.3, \quad (5)$$

$$P(U_i = 1|\theta_i) = 0.35 \text{ y } P(U_i = 1|\theta_h) = 0.65, \text{ si } 0.3 \leq c_i < 0.35, \quad (6)$$

$$P(U_i = 1|\theta_i) = 0.5 - \varepsilon \text{ y } P(U_i = 1|\theta_h) = 0.8 - \varepsilon, \text{ si } 0.35 \leq c_i < 0.5, \text{ con } \varepsilon > 0, \quad (7)$$

Se nota además que ninguna pareja (θ_i, θ_h) cumple los tres criterios si $c_i \geq 0.5$. Sin embargo, conviene recordar que cuando el parámetro c_i es muy grande se suele considerar que el ítem presenta poca utilidad en la construcción de la escala. Una vez determinados los intervalos $[\theta_i, \theta_h]$ para cada ítem, el objetivo es encontrar el conjunto de θ_p , $p = 0, 1, \dots, P$, tal que sea posible anclar a los ítems en P niveles, con $1 < P < I$. El procedimiento para encontrarlos consta de los siguientes pasos:

1. Ordenar los I ítems ascendentemente de acuerdo con el θ_i hallado para cada uno de ellos.
2. Tomar como nivel θ_0 el menor de estos θ_i .
3. Si $p = 1$, tomar como candidato para θ_1 , llamémoslo θ_1^* , al θ_h del ítem que produjo θ_0 . Si $p > 1$, tomar al θ_h del ítem siguiente al que produjo θ_{p-1} como candidato para θ_p .
4. Comparar el candidato con el intervalo hallado para el siguiente ítem:
 - a. Si $\theta_i^1 < \theta_p^* < \theta_h$, se procede a tomar el θ_h de este ítem como nuevo candidato y se repite la comparación con el siguiente intervalo.
 - b. Si $\theta_p^* \geq \theta_h$, se conserva θ_i^* como candidato y se repite la comparación con el siguiente intervalo.
 - c. Si $\theta_p^* \leq \theta_i$, entonces θ_p^* determina el nivel de desempeño θ_p .
5. Repetir los pasos 3 y 4 para determinar θ_p para los niveles $p = 2, \dots, P$ hasta haber comparado los I ítems.

Es importante aclarar que este procedimiento se propone como una herramienta para facilitar la descripción de la escala de habilidades al permitir encontrar puntos en la misma que definan niveles de desempeño, como fueron propuestos por Beaton y Allen (1992); así como aquel procedimiento puede ser incapaz de anclar todos los ítems que componen una prueba, el procedimiento aquí propuesto puede no lograr encontrar un número de niveles de desempeño con $P > 1$ a partir de todos los ítems de la prueba.

En la Figura 2 se representan los intervalos de dos conjuntos de ítems ordenados de acuerdo con el valor de θ_i encontrado para cada uno de ellos; en esta figura se puede ver claramente que el procedimiento propuesto es capaz de identificar los sitios de la gráfica donde se presentarían “saltos” entre los intervalos de anclaje de los ítems, y que si no se presentan, el procedimiento no puede identificar varios niveles.

En aquellos casos en que el procedimiento propuesto falla en encontrar $P > 1$ niveles de desempeño a partir del conjunto completo de ítems, es posible eliminar algunos ítems de la descripción por niveles de desempeño. Este proceso no puede ser automatizado y en su realización debe recordarse que estos ítems únicamente se ignoran en la construcción de descripciones de las capacidades o de las características de los individuos cuya habilidad se encuentra por encima de alguno de los niveles de desempeño encontrados.

Estas interpretaciones de los ítems anclados en distintos niveles buscan dar generalizaciones sobre las habilidades de la población de examinados, así como a la hipotética población de ítems que satisfacen los criterios para el respectivo nivel de desempeño; por consiguiente, la interpretación de los ítems debería, en principio, ser realizada por especialistas en el área específica que la escala pretende evaluar. Las

generalizaciones que ellos dan a los niveles de desempeño permiten, por una parte, generar teoría sobre la relación entre la habilidad de los examinados y los ítems que componen la prueba, y por otra, resumir las conclusiones sobre los resultados de la escala para un público poco técnico en el área (Beaton & Allen, 1992).

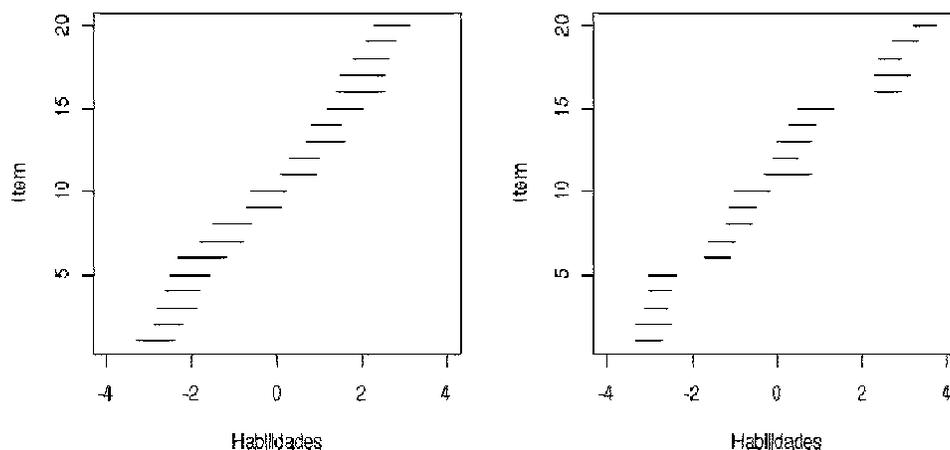


Figura 2. Aproximación gráfica a los niveles de desempeño en dos conjuntos de ítems.

Aplicación: ‘Personas en mi vida’

A continuación se presenta el análisis de una escala del cuestionario ‘Personas en mi vida’ empleando modelos IRT 3-PL y utilizando el procedimiento de anclaje de la escala para interpretar los resultados. Éste es un cuestionario diseñado por Cook et al. (1995) para evaluar las percepciones de los niños sobre sus relaciones afectivas con los padres, maestros y pares en la edad escolar; se deriva del IPPA (Inventario de Apego con Padres y Pares), diseñado por Armsden y Greenberg (1987), el cual evalúa las percepciones de los adolescentes sobre sus relaciones afectivas con padres y pares.

El cuestionario ‘Personas en mi vida’ está compuesto por 78 ítems distribuidos en cinco partes: Relaciones con padres, relaciones con pares, relaciones con profesores, relaciones con la escuela y relaciones con el vecindario. Cada una de estas partes contiene sus respectivas escalas. La primera parte se compone de tres escalas: Confianza, comunicación y alienación; la segunda la componen confianza, comunicación, alienación y delincuencia; la tercera se integra por afiliación y alienación respecto a los profesores; la cuarta por afiliación y peligrosidad escolar; y la última está integrada por las escalas de barrio positivo y peligrosidad del barrio. Éste es un instrumento de autorreporte específico para niños escolarizados entre 9 y 12 años. Cada pregunta es respondida en una escala Likert ordenada de 1 a 4 (‘Nunca’ a ‘Siempre’) de acuerdo con la frecuencia con que el niño o niña evalúa que le sucede la situación descrita por el ítem.

El presente análisis se realizó a partir de una adaptación del cuestionario al español aplicada en niños y niñas de Bogotá por Camargo y Mejía (2005), y analizada desde la Teoría Clásica de los Tests (TCT o CTT por su sigla en inglés) (Camargo et al., 2007).

Participantes

Se trabajó con una muestra no aleatoria de 2274 niños (49.2% niñas y 50.8% niños) entre los 8 y 14 años de edad ($M = 10.84$, $D.E. = 1.17$), quienes se encontraban estudiando en instituciones educativas bogotanas: 6 públicas y 4 privadas. La muestra se distribuyó de acuerdo con las características

socioeconómicas de Bogotá, de manera que el 47.5% de los niños pertenecía al estrato bajo (estratos 1 y 2), el 50.7% al estrato medio (estratos 3 y 4) y el 1.6% al estrato alto (estratos 5 y 6).

Procedimiento

Se analizaron las respuestas de los 2274 niños a las 78 preguntas que conforman el cuestionario ‘Personas en mi vida’. Para cada una de las partes que componen la prueba, se evaluó la dimensionalidad del conjunto de ítems que la componen, teniendo en cuenta la estructura de escalas propuestas por los autores del instrumento y habiendo calificado las preguntas de forma dicótoma (0 si el niño respondió que ‘Nunca’ o ‘Algunas veces’ le ocurría la situación descrita, y 1 si el niño respondió que ‘Muchas veces’ o ‘Siempre’). En todos los casos la estructura teórica y la hallada en la muestra coinciden, excepto las escalas de comunicación y confianza de las partes correspondientes a relaciones con padres y con pares que aparecen reunidas en un solo factor (Camargo & Mejía, 2005).

Se tomó una de las escalas de la prueba, la cual recibió el nombre de Afiliación, compuesta por ítems de Comunicación y Confianza de la parte de Relaciones con Padres. Esta escala fue analizada con un modelo IRT 3-PL, empleando el paquete *ltm* (Rizopoulos, 2006) del programa estadístico R (R Development Core Team, 2004). En el análisis de esta prueba es claro que el parámetro c no puede ser interpretado en el sentido de pseudoazar o adivinación, puesto que no hay respuestas correctas o incorrectas; en este caso, la interpretación más apropiada para este parámetro sería una tendencia a responder la pregunta por deseabilidad social, es decir, a indicar que se está de acuerdo con lo descrito en el reactivo aun cuando no describe la realidad del sujeto, con el fin de presentar una imagen favorable de sí mismo ante la persona que aplica la prueba o el receptor de los resultados (Brown, 1980).

Resultados

En la Tabla 1 se presentan las estimaciones del modelo 3-PL correspondiente a la escala de Afiliación de la sección de Relaciones con Padres del cuestionario. Los parámetros de los individuos presentaron una media de -0.1285 y una desviación estándar de 0.824; la distribución de sus valores se encontró sesgada hacia la izquierda, con un coeficiente de asimetría de -0.476 (Mínimo=-2.762, Máximo=1.012).

De acuerdo con el procedimiento propuesto anteriormente, se hallan los intervalos $[\theta_i, \theta_n]$, para cada ítem. Estos intervalos se presentan en la Tabla 1 y se pueden ver ilustrados en la Figura 3 ordenados de acuerdo con el valor de θ_i . Junto a cada intervalo aparece el número del ítem correspondiente.

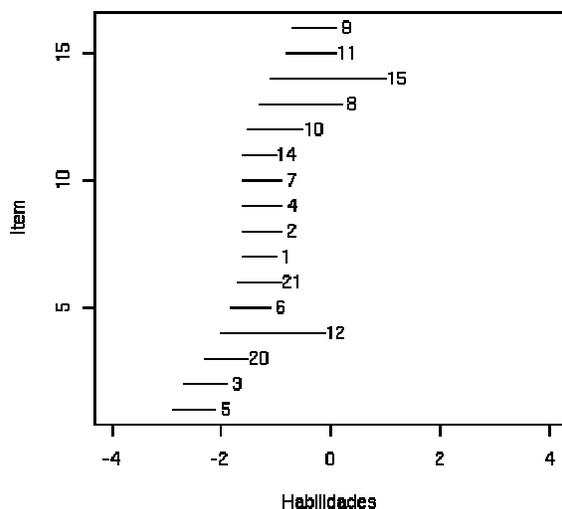


Figura 3. Afiliación: intervalos de anclaje.

En la Figura 3 puede verse que el procedimiento no produce un número de niveles de desempeño mayor que 1, pero que si se eliminan los ítems 8, 10, 12, 15 y 20 es posible obtener una clasificación en tres grupos de preguntas. De este modo se obtienen los siguientes niveles $(\theta_0, \theta_1, \theta_2, \theta_3) = (-2.9, -1.9, -0.7, 1)$, donde los ítems 3 y 5 pertenecen al nivel 1, los ítems 1, 2, 4, 6, 7, 10, 14 y 21 pertenecen al nivel 2, los ítems 9 y 11 pertenecen al nivel 3, y los demás ítems no quedan anclados en ningún nivel.

Teniendo en cuenta los ítems anclados a cada nivel, se puede indicar que el nivel 1 caracteriza un grado de confianza y comunicación con los padres tal que se percibe que la relación con ellos es

positiva, aunque no necesariamente cercana; en cuanto al nivel 2, el grado es tal que se percibe que los padres muestran un verdadero interés por su hijo; en el nivel 3, el grado es tal que el niño percibe una relación aún más cercana, y reporta que puede acudir a sus padres en busca de apoyo emocional.

En la muestra de niños a quienes se aplicó el cuestionario, el 13.63% está en el nivel 3, el 68.62% está en el nivel 2, el 15.16% está en el nivel 1, el 2.59% está en el nivel 0 y ninguno está por debajo del nivel 0. De los niños en cada nivel se espera que perciban su relación con sus padres como descrita por aquello que caracteriza el nivel, así como por aquello que caracteriza los niveles menores. En cuanto a los niños que se ubican en el nivel 0, se puede esperar que no perciban ninguna de las características descritas en los otros niveles en su relación con sus padres.

Tabla 1.
Parámetros estimados e intervalos de anclaje para el factor de Afiliación (Comunicación y confianza) con los padres.

Ítem	Discriminación	Dificultad	Pseudo-azar	θ_1	θ_2
1	1.937	-1.305	0.005	-1.6	-1.0
2	1.731	-1.23	0.009	-1.6	-0.9
3	1.473	-2.289	0.025	-2.7	-1.9
4	1.742	-1.213	0.002	-1.6	-0.9
5	1.753	-2.502	0.008	-2.9	-2.1
6	1.803	-1.451	0.009	-1.8	-1.1
7	1.931	-1.259	0.104	-1.6	-0.9
8	1.54	-0.26	0.398	-1.3	0.2
9	1.725	-0.296	0.094	-0.7	0.1
10	1.241	-1.036	0.019	-1.5	-0.5
11	1.465	-0.365	0.025	-0.8	0.1
12	1.12	-0.776	0.371	-2.0	-0.1
14	1.948	-1.303	0.001	-1.6	-1.0
15	0.587	-0.037	0.007	-1.1	1.0
20	1.665	-1.901	0.001	-2.3	-1.5
21	1.64	-1.305	0.001	-1.7	-0.9

Discusión

En este trabajo se han presentado las bases de los principales modelos de la Teoría de Respuesta al Ítem y la perspectiva del anclaje de la escala en la interpretación de las pruebas; dentro de esta perspectiva se presentó una propuesta para la obtención empírica de los niveles de desempeño.

El procedimiento propuesto para la obtención empírica de niveles de desempeño permite realizar una aproximación más ajustada del atributo representado en el instrumento de medición a partir de los reactivos que efectivamente lo componen. En primer lugar, permite conseguir los dos propósitos principales que busca el anclaje de la escala; esto es, al ubicar niveles de desempeño tales que un conjunto de ítems pertenece a cada uno de estos niveles, facilita la generación de teoría sobre la relación entre el atributo evaluado en los individuos y los ítems que componen la prueba, y facilita la comunicación de los resultados de la evaluación a un público que no es experto en el área. En segundo lugar, el procedimiento propuesto garantiza que los niveles de desempeño elegidos sean interpretables, puesto que ningún nivel de desempeño –diferente al nivel 0– queda sin ítems que pertenezcan al mismo, y que para un conjunto específico de ítems, los niveles elegidos son aquellos que bajo los criterios de anclaje mejor pueden describir a los ítems aplicados.

Es importante recalcar que esta propuesta se ubica en el eslabón de interpretación de la escala obtenida en la cadena de pasos que se realizan al analizar un instrumento de evaluación con los modelos TRI, y por consiguiente, requiere que el instrumento haya sido trabajado de forma adecuada en la construcción, estimación, calibración y elección de los reactivos que lo componen. Así mismo, debe recordarse que esta propuesta es sólo una herramienta de apoyo en la descripción e interpretación de la escala, un complemento de esta descripción, no un sustituto, y por sí sola no constituye un anclaje, ni una interpretación, ni sustituye el trabajo de los especialistas en el área para darle sentido a la escala.

Mediante la aplicación presentada se puede apreciar que el procedimiento propuesto resulta ser una herramienta promisoría en la interpretación de las escalas en instrumentos construidos bajo los modelos TRI. Su uso mostró una interpretación coherente con la clasificación teórica de los estilos de apego (Ainsworth, Blehar, Waters & Wall, 1978), mediante la cual se obtuvo una mejor comprensión de esta escala compuesta por preguntas de las escalas de Comunicación y Confianza en el instrumento original. En este aspecto, los diferentes niveles de desempeño que se pudieron identificar en la escala permiten dar cuenta de las características propias de la clasificación del apego dada por Ainsworth et al. (1978). El análisis de todas las escalas del instrumento ‘Personas en mi vida’ supera los objetivos de este trabajo, pero es importante para la completa comprensión del mismo y para la descripción de las relaciones afectivas de los niños bogotanos a quienes se aplicó; éste es un trabajo en progreso y será publicado posteriormente.

Finalmente, el procedimiento tiene la ventaja práctica de que puede ser casi completamente automatizado. Es posible solicitar al primer autor las funciones para el paquete R (R Development Core Team, 2004) para la realización de este procedimiento.

Referencias

- Ainsworth, M., Blehar, M. C., Waters, E. & Wall, S. (1978). *Patterns of attachment*. Hillsdale, New Jersey, Estados Unidos: Lawrence Erlbaum Associates.
- Andrade, D. F. d., Tavares, H. R. & Valle, R. d. C. (2000). Teoria da resposta ao item: Conceitos e aplicações. Recuperado el 8 de diciembre de 2006 de http://www.inf.ufsc.br/~dandrade/TRI/LivroTRI_pdf.zip
- Armsden, G. C. & Greenberg, M. T. (1987). The inventory of parent and peer attachment: Individual differences and their relationship to psychological well-being in adolescence. *Journal of Youth and Adolescence*, 16(5), 427–454.
- Baker, F. B. & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques*. (2nd ed.). Nueva York: Marcel Dekker, Inc.
- Bartholomew, D. J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society*, 42(3), 293–321.
- Bartholomew, D. J. (1984). The foundations of factor analysis. *Biometrika*, 71(2), 221–232.
- Beaton, A. E. & Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics*, 17(2), 191–204.
- Beaton, A. E. & Johnson, E. G. (1992). Overview of the scaling methodology used in the national assessment. *Journal of Educational Measurement*, 29(2), 163–175.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee’s ability. En F. M. Lord & M. R. Novick (Eds.). *Statistical theories of mental test scores* (pp. 395–479). Reading, Massachusetts: Addison-Wesley.
- Brown, F. G. (1980). *Principios de la medición en psicología y educación*. México: Manual Moderno.
- Camargo, S. L. & Mejía, G. L. (2005). *Validación y estandarización del cuestionario “Personas en mi vida” en niños y niñas bogotanos entre 9 y 12 años de edad*. Tesis de pregrado no publicada, Universidad Nacional de Colombia, Colombia.
- Camargo, S. L., Mejía, G. L., Herrera, A. N. & Carrillo, S. (2007). Adaptación del cuestionario “Personas en mi vida” en niños y niñas bogotanos entre 9 y 12 años de edad. *Acta Colombiana de Psicología*, 10(2), 83-93.
- Cook, E., Greenberg, M. & Kusche, C. (1995, marzo). *People in my life: Attachment relationships in middle childhood*. Paper presented at the Society for Research in Child Development; Indianapolis, IN, March. 1995.

- Ferrando, P. J. (1996). Relaciones entre el análisis factorial y la teoría de respuesta a los ítems. En J. Muñiz (Ed.). *Psicometría* (pp. 555–612). Barcelona, España: Universitat.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications.
- Lazarsfeld, P. F. (1955). Recent developments in latent structure analysis. *Sociometry*, 18(4), 391–403.
- Lord, F. M. (1950). A theory of test scores. *Psychometric Monographs*, 7. Chicago: University of Chicago Press.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Lord, F. M. (1986). Maximum likelihood and bayesian parameter estimation in item response theory. *Journal of Educational Measurement*, 23(2), 157–162.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley.
- Mislevy, R. & Bock, D., R. (1997). *BILOG 3 for windows* [Manual y software de cómputo]. Chicago, Estados Unidos: Scientific Software Inc.
- Muñiz, J. (1996). *Psicometría*. Barcelona, España: Universitat.
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Ediciones Pirámide.
- Pinto, M. (2006). *Teoría de respuesta al ítem: Estimación y modelación*. Tesis de pregrado no publicada, Universidad Nacional de Colombia.
- R Development Core Team. (2004). *R: A language and environment for statistical computing* [Manual y software de cómputo]. Vienna, Austria.
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, 17(5). Recuperado de <http://www.jstatsoft.org/v17/i05/paper>
- Van der Linden, W. J. & Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. En W. J. van der Linden & R. K. Hambleton (Eds.). *Handbook of modern item response theory* (pp. 1-28). New York: Springer-Verlag.
- Wright, B. D. & Stone, M. H. (1998). *Diseño de mejores pruebas utilizando la técnica de Rasch*. México: CENEVAL. (Trabajo original publicado en 1979).

Manuscrito recibido en Enero de 2008
Aceptado para publicación en Junio de 2008