

# REVISIÓN DE INVESTIGACIONES RECIENTES SOBRE LA APLICACIÓN DE LA TEORÍA DE RESPUESTA AL ÍTEM AL CHILD BEHAVIOR CHECKLIST

## REVISION OF RECENT INVESTIGATIONS ABOUT THE APPLICATION OF ITEM RESPONSE THEORY TO THE CHILD BEHAVIOR CHECKLIST

*Abal, Facundo Juan Pablo<sup>1</sup>; Lozzia, Gabriela Susana<sup>2</sup>; Blum, G. Diego<sup>3</sup>;  
Aguerri, María Ester<sup>4</sup>; Galibert, María Silvia<sup>5</sup>; Attorresi, Horacio Félix<sup>6</sup>*

### RESUMEN

El Child Behavior Checklist (CBCL) es un formulario que permite registrar problemas comportamentales y competencias sociales de niños y adolescentes. Si bien se construyó desde el enfoque clásico de la Teoría de los Test, investigaciones recientes mostraron la utilidad de aplicar la Teoría de Respuesta al Ítem (TRI) para modelizar las variables del CBCL. El desarrollo de este trabajo revisa las consideraciones que se vieron obligados a tomar los respectivos autores para cumplir con las exigencias de los modelos de la TRI. Se describe la información que brindan las Curvas Características de los Ítems y cómo su análisis podría contribuir para optimizar el tiempo de administración del CBCL. Asimismo, se sintetizan los resultados alcanzados en relación con el estudio de la invarianza de las medidas obtenidas a través de adaptaciones transculturales de este formulario.

### Palabras clave:

Child Behavior Checklist - Teoría de Respuesta al Ítem - Curva Característica del ítem

### ABSTRACT

The Child Behavior Checklist (CBCL) is a questionnaire that allows the assessment of behavioral problems and social competences of children and adolescents. Although it was developed on the basis of the classical approach of the Test Theory, recent investigations have revealed the usefulness of applying the Item Response Theory (IRT) to model the variables of CBCL. This paper revises the considerations made by the respective authors in order to meet the demands of the IRT models. The information provided by the Characteristic Curves of Items is described as well as the manner in which the analysis thereof might contribute to optimize the administration period of CBCL. Furthermore, the invariance of the measures obtained through transcultural adaptations of this questionnaire is analyzed and the results are summarized.

### Key words:

Child Behavior Checklist - Item Response Theory - Item Characteristic Curve

<sup>1</sup>Licenciado en Psicología. Ayudante de Primera Interino de la Cátedra II de Estadística de la Facultad de Psicología de la Universidad de Buenos Aires. Becario de Doctorado del Proyecto UBACyT P043. Investigador en el Proyecto de Investigación PICT 20909 de la Agencia (ANPCyT). E-mail: fabal@psi.uba.ar

<sup>2</sup>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. Becaria de Doctorado de CONICET. Investigadora en los Proyectos P043 de la Universidad de Buenos Aires y PICT 20909 de la Agencia (ANPCyT). E-mail: glozzia@psi.uba.ar

<sup>3</sup>Especialista en Metodología de la Investigación Científica. Licenciado en Psicología. Ayudante de Primera Interino de la Cátedra I de Metodología de la Investigación de la Facultad de Psicología, U.B.A. Becario de Doctorado e investigador del Proyecto UBACyT P043. E-mail: blumworx@gmail.com

<sup>4</sup>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. Codirectora del Proyecto P043 de la Universidad de Buenos Aires e Investigadora del Proyecto PICT 20909 de la Agencia (ANPCyT). E-mail: maguerri@psi.uba.ar

<sup>5</sup>Idem 4 E-mail: galibert@psi.uba.ar

<sup>6</sup>Licenciado en Ciencias Matemáticas. Profesor Regular Titular de la Cátedra II de Estadística de la Facultad de Psicología de la Universidad de Buenos Aires. Director de los Proyectos con Subsidio P043 de la Universidad de Buenos Aires y PICT 20909 de la Agencia (ANPCyT). Director de Becarios de UBACyT, CONICET y ANPCyT. E-mail: hatorre@psi.uba.ar

### Revisión de investigaciones recientes sobre la aplicación de la Teoría de Respuesta al Ítem al Child Behavior Checklist

El cuestionario Child Behavior Checklist (CBCL) fue el primer formulario elaborado por Achenbach (Achenbach & Edelbrock, 1983) para registrar problemas comportamentales y competencias sociales de niños y adolescentes con edades comprendidas entre 4 y 18 años. Si bien su aplicación es útil en el contexto clínico, este instrumento es comúnmente utilizado como herramienta de cribado en estudios epidemiológicos. Por consiguiente, su uso en el ámbito de la prevención es valorado y reconocido internacionalmente para la detección de casos psiquiátricos en población infantojuvenil (Navarro, Doménech, de la Osa & Ezpeleta, 1998).

El CBCL permite la evaluación de los problemas comportamentales y emocionales de los niños desde un enfoque dimensional del diagnóstico psicopatológico (Samaniego, 2008). Su construcción se basó en la implementación de técnicas de estadística multivariada para el análisis de los datos recolectados con este formulario. De esta manera, se pudo identificar un conjunto de dimensiones latentes considerando la covariación de los problemas de conducta o síntomas reportados. Estas dimensiones fueron descritas como síndromes o desórdenes y cada una de ellas conforman un continuo en el que es posible ubicar a todos los individuos según la severidad que con que se manifiesta la sintomatología (Cervone et al., 2003).

Unos años después de la aparición del CBCL se desarrollaron dos formas, una para ser respondida por los docentes (Teacher's Report Form, TRF) a cargo del niño (Achenbach, 1991a) y otra versión autorreporte (Youth Self Report, YSR) para jóvenes (Achenbach, 1991b). La denominación CBCL quedó actualmente restringida para la versión que responden los padres. Así también, aunque la primera versión del CBCL incluía los grupos de 4 y 5 años, las revisiones posteriores discriminaron formas distintas para niños en edad pre-escolar (1.5 a 5 años) y en edad escolar (6 a 18 años). En la actualidad, todas las formas derivadas del CBCL original componen un sistema de evaluación multi-informante denominado Achenbach System of Empirically Based (ASEBA; Achenbach & Rescorla, 2001).

En numerosos análisis factoriales, tanto exploratorios como confirmatorios de los datos obtenidos a partir de las distintas formas del CBCL (versiones para padres, profesores y autorreporte), dieron como resultado un mismo modelo jerárquico de dos niveles que ha sido replicado en sucesivas revisiones (Ngo, 2006). El modelo permite describir el comportamiento de los evaluados mediante un conjunto de ocho escalas estrechas de síndromes surgidas de un análisis factorial de primer orden (Retraimiento, Quejas somáticas, Ansioso-Depresivo, Problemas de pensamiento, Problemas de atención, Problemas sociales, Agresividad y Conducta

antisocial). Las primeras tres escalas conforman una dimensión de segundo orden (Internalizante) y las dos últimas se suman para formar otro factor denominado Externalizante. El resto de las dimensiones de primer orden tienen un pesaje equivalente en las dimensiones Internalizante y Externalizante.

El CBCL original, sus posteriores modificaciones y las distintas formas fueron adaptados y utilizados en investigaciones de diversos países (Navarro et al., 1998; Ngo, 2006; Samaniego, 2008). En uno de los estudios transculturales más recientes - que demuestra la robustez del modelo de Achenbach - Ivanova et al. (2007) corroboraron en 23 culturas la taxonomía de ocho síndromes derivados de la YSR mediante un análisis factorial confirmatorio. En Argentina, Samaniego (2008) validó una versión para padres del CBCL en su sección de problemas comportamentales. Se recolectaron evidencias de validez y confiabilidad satisfactorias para nuestro medio, lo que permitió su aplicación para el estudio de la prevalencia de trastornos psíquicos en niños escolarizados de 6 a 11 años del partido de San Isidro (Samaniego, 2004). Así también, dado que se estudió la equivalencia de contenido, semántica y técnica de la adaptación, se pudo efectuar una comparación de esta prevalencia con resultados obtenidos en otros países.

El factor común de las investigaciones referidas hasta aquí es que la teoría psicométrica utilizada para modelar las respuestas al CBCL ha sido la Teoría Clásica de los Tests (TCT). A pesar de su extendida aplicación, es ampliamente reconocido que el análisis clásico resulta útil pero no del todo exhaustivo. La principal limitación de la TCT es que la medida de una variable no es completamente independiente del grupo normativo y de los elementos que componen el test utilizado. Esto plantea problemas al momento de interpretar los resultados y, principalmente, cuando se trata de establecer comparaciones entre los puntajes obtenidos en diferentes culturas.

El objetivo de este estudio es revisar algunas investigaciones recientes que modelizan las respuestas al CBCL desde una perspectiva psicométrica superadora de la TCT: la Teoría de Respuesta al Ítem. Específicamente, los artículos encontrados y que serán considerados para el presente trabajo, utilizaron la versión para padres (Ngo, 2006) y la versión autorreporte (YSR) para jóvenes de 11 a 18 años (Curran, et. al. 2008; Lambert et al., 2003; Lambert, Essau, Schmitt & Samms-Vaughan, 2007).

### DESARROLLO

La *Teoría de Respuesta al Ítem* (TRI) engloba un conjunto de modelos probabilísticos que son utilizados para formalizar la relación existente entre la respuesta que un sujeto da a un ítem y la dimensión latente ( $\theta$ ) que explica esta conducta. Estos modelos permiten vincular mediante una función matemática la probabilidad de dar una determinada respuesta al ítem para cada nivel de la

variable medida por éste. El gráfico de dicha función matemática se denomina *Curva Característica del Ítem* (CCI). Las CCIs correspondientes a cada uno de los ítems de la prueba serán diferentes y su forma específica quedará determinada cuando se estimen los parámetros correspondientes al modelo elegido.

A partir de los patrones de respuestas observados en una muestra, la TRI estima de forma conjunta los parámetros de la curva de cada ítem y el valor del rasgo para cada sujeto. Las CCI quedan caracterizadas por sus propios parámetros independientemente de cómo se distribuya el rasgo latente en la población en la que es administrado el test y de cuáles sean los parámetros de los demás ítems. Además, la medida de rasgo de una persona no depende ni de los parámetros de los ítems ni de la medida de los demás individuos. Son estas características básicas de la TRI las que ayudan a encontrar respuestas a los principales inconvenientes observados en la TCT.

Aunque la aplicación de la TRI es extremadamente infrecuente en ámbito de la salud y la prevención, en los últimos años se observó un creciente interés por parte de los investigadores del área (e.g. Kessler & Mroczek, 1994; Reiser, 1989; Stansbury, Ried & Velozo, 2006; Zwinderman, 1990). En relación con su aplicación al formulario CBCL, Lambert et al. (2007) reconocieron que recién actualmente se están realizando los primeros esfuerzos para obtener evidencias de validez y confiabilidad con indicadores de la TRI.

### **CONSIDERACIONES PREVIAS A LA APLICACIÓN**

La aplicación de los modelos de la TRI a las variables del CBCL ha requerido a los investigadores tener en cuenta un conjunto de consideraciones previas. Como estas consideraciones condicionan la aplicación de la TRI, se vieron obligados a tomar ciertas decisiones respecto del tratamiento de los datos. Si bien existe una diversidad de modelos para describir las respuestas a los ítems, los más desarrollados al momento son modelos unidimensionales para datos dicotómicos, lo cual entró en conflicto con la estructura jerárquica de síndromes del modelo de Achenbach y el formato de respuesta politómica que tienen los ítems del CBCL.

El supuesto de unidimensionalidad requerido por los modelos de la TRI tiene una enorme implicancia en el análisis de los datos obtenidos con el CBCL. Lambert et al. (2007) decidieron trabajar sobre las dimensiones Internalizante y Externalizante. Estos investigadores obtuvieron una solución bifactorial y analizaron cada conjunto de ítems como unidimensionales. Las escalas resultantes son similares a las escalas ampliadas de Achenbach pero también incorporan ítems correspondientes a otros factores de primer orden no contemplados en la escala original. Ngo (2006) en cambio, consideró como unidimensional a cada una de las escalas estrechas. Por último, Curran et al. (2008) se centraron

en la medición de sintomatología internalizante. Estos autores sólo tomaron ítems correspondientes a esa escala y los combinaron con indicadores de otros formularios que miden constructos similares.

Como ya se mencionó, otra importante decisión metodológica involucra a la elección del modelo ajustado en función de la cantidad de opciones de respuesta del ítem. La TRI considera cada ítem como una unidad de análisis para formalizar la relación entre la variable medida y la respuesta observada. Por lo tanto, cobra significativa importancia el tipo de respuesta que se obtendrá del individuo en virtud de que constituye el único dato disponible para establecer dicha relación. Los ítems del YSR y del CBCL son respondidos en una escala graduada de tres puntos. Las etiquetas asignadas a estas opciones remiten a la frecuencia de aparición de los problemas referidos por cada ítem: *nunca observado, algo o algunas veces cierto y muy cierto o a menudo cierto*. Aunque desde la perspectiva clásica estas respuestas son codificadas con 0, 1 y 2 respectivamente, sólo Ngo (2006) mantuvo este formato de puntuación. Por ello modelizó las respuestas a los ítems mediante el modelo politómico de respuesta ordenada de Samejima. En el resto de los trabajos, los autores dicotomizaron las opciones de respuesta considerando la ausencia o presencia del indicador. En consecuencia, estas aplicaciones se efectuaron con Modelos Logísticos de 1, 2 y 3 parámetros. A sabiendas de que la dicotomización implica una pérdida de la información contenida en los ítems, Curran et al. (2008) justificaron la fusión de las dos categorías más altas con la baja frecuencia de elección que presenta la última opción de respuesta.

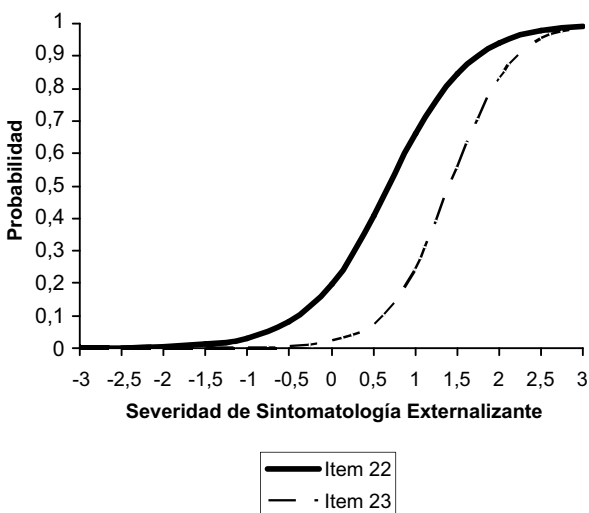
### **ANÁLISIS DE LAS CURVAS CARACTERÍSTICAS DE LOS ÍTEMS**

El análisis de las CCI y sus parámetros depende totalmente del modelo ajustado. En virtud de la variedad de los modelos utilizados por los distintos autores, se consideró pertinente abordar este apartado considerando únicamente, a modo de ejemplo, el estudio de Lambert et al (2007). La interpretación de los resultados permite dos niveles de análisis. Uno más minucioso que describe las características del ítem y otro más amplio que contempla las propiedades psicométricas de todos los ítems que evalúan la dimensión.

Lambert et al. (2007) administraron el YSR a una muestra amplia de adolescentes jamaquinos y modelizaron las respuestas de las dimensiones Internalizante y Externalizante con el Modelo Logístico de 2 parámetros (ML2p). En la Figura 1 se observan las curvas características de dos ítems de la escala Externalizante graficadas a partir de los resultados publicados en Lambert et al. (2007). El ítem 22 se refiere a la manifestación de desobediencia en el hogar mientras que el ítem 23 concierne a desobedecer en el contexto escolar. Como se puede apreciar, cuanto más alto es el nivel de severidad

del síntoma percibido por el adolescente, mayor probabilidad tendrá de dar una respuesta afirmativa al ítem. Pero la respuesta del individuo a un ítem no sólo se debe a su nivel en la escala sino que también depende de las propiedades psicométricas del reactivo (i.e. los parámetros). El ML2p describe el funcionamiento del ítem mediante un parámetro de localización ( $b$ ) y otro de discriminación ( $a$ ). El parámetro de localización es un punto ubicado en la escala de la variable que define la transición entre la probabilidad de reportar ausencia del problema y la de considerarlo como presente en el individuo examinado. El parámetro de discriminación refleja en qué medida el indicador diferencia a los evaluados con un nivel de severidad por encima o por debajo del parámetro de localización.

Figura 1. Comparación de la Curvas Características de dos Ítems.



Sabiendo que los parámetros estimados para el ítem 22 son  $a_{22} = 1.21$  y  $b_{22} = 0.68$  y que los parámetros del ítem 23 son  $a_{23} = 1.62$  y  $b_{23} = 1.42$  es posible efectuar un análisis comparativo del aporte de estos indicadores a la medición de la dimensión Externalizante. En relación con el parámetro  $b$  se observa que es necesario percibir un nivel de sintomatología Externalizante más alto para reportar una conducta de desobediencia en la escuela que en el hogar ( $b_{22} < b_{23}$ ), lo cual puede ser esperable desde un punto de vista clínico. Respecto de los índices  $a$ , aunque el ítem 23 discrimina mejor que el ítem 22 ( $a_{23} > a_{22}$ ) la diferencia es prácticamente imperceptible en la Figura 1. La interpretación adecuada de estas curvas muestra que el ítem 23 tiene un mayor poder discriminativo en torno a los niveles de severidad medio-altos ( $b_{23} = 1.42$ ), mientras que en el ítem 22 la capacidad discriminativa se ubica en torno a los valores de severidad intermedios ( $b_{22} = 0.68$ ). Por lo tanto, cada uno de ellos aporta información en diferentes niveles de la escala. Como es posible observar, la modelización de las res-

puestas mediante la TRI también ofrece una valiosa información para comprender el constructo evaluado. Su aplicación devela la relación que existe entre los síntomas propuestos por los ítems y la capacidad de éstos para discriminar en un valor específico del espectro Externalizante. La respuesta afirmativa a un ítem con un parámetro de localización elevado muestra una alta probabilidad de que el sujeto se ubique entre los valores más altos de severidad. Es decir, a partir de los parámetros de localización es posible realizar un escalamiento de los ítems indicando cuáles discriminan en niveles de severidad más bajos o en niveles más altos.

Ahora bien, la TRI además permite realizar un análisis más global de los resultados a partir de la interpretación de los parámetros de todos los ítems administrados. Por ejemplo, Lambert et al. (2007) hallaron que todos los parámetros de localización de los ítems que cargaban en la dimensión Externalizante se concentraban en los niveles medio-altos de la variable. Esto resulta consistente si se piensa que el objetivo de la escala es detectar con mayor precisión los niveles altos de sintomatología Externalizante. Sin embargo, Lambert et al (2007) también encontraron que la mayoría de los parámetros  $b$  de los ítems de la escala Internalizante se agrupaban en niveles medio-bajos de severidad. Esto reveló una debilidad en las características psicométricas de la escala Internalizante dado que demostró que la misma carecía de elementos para discriminar, en la parte del espectro que evalúa, niveles elevados de severidad sintomática.

La relevancia de conocer las propiedades psicométricas de los ítems puede trasladarse al campo de la investigación epidemiológica aplicada. Pedreira et al. (1993) remarcaron como un defecto la extensión de las escalas de cribado como el CBCL y la necesidad de acortarlas considerando los ítems más discriminativos. En uno de sus estudios, Lambert et al. (2003) hallaron que 3 de cada 4 ítems de todas las escalas estrechas presentaban bajos índices de discriminación. Los autores afirmaron que, como los mismos aportan poca información, su eliminación podría resultar conveniente para obtener una reducción significativa del tiempo de administración.

Un procedimiento acorde con el objetivo de detectar individuos con psicopatología en un menor tiempo se podría llevar a cabo administrando exclusivamente los ítems que muestran mayor poder discriminativo (parámetro  $a$ ) en torno al valor definido como punto de corte. La intervención podría instrumentarse rápidamente luego de una administración de pocos ítems, pero altamente discriminativos.

Otra forma de obtener una medición eficiente en el marco de la TRI resulta de la implementación de un test adaptativo informatizado. En este caso, un software selecciona progresivamente los ítems más apropiados para la medición de una persona en función del nivel de severidad sintomática que va manifestando en cada respuesta.

### **ESTUDIO DE LA INVARIANCIA DE LOS PARÁMETROS DE LOS MODELOS APLICADOS**

Si bien las investigaciones revisadas aplican algún modelo de la TRI, el fin último de estos estudios es recolectar evidencias que corroboren la invarianza o no de las medidas a través de las distintas culturas y/o países. En la TRI, así como se estiman los parámetros de un modelo a partir de una muestra de individuos, también se espera que la estimación surgida de otro conjunto de datos alcance parámetros similares (salvo por fluctuaciones del azar). En efecto, si la respuesta a un ítem sólo depende del nivel de la variable que se desea medir, dos adolescentes con el mismo nivel en la dimensión latente deberían tener la misma probabilidad de reportar un síntoma independientemente de la población de pertenencia. Cuando esta suposición no se verifica se afirma que el ítem presenta un funcionamiento diferencial (Differential Item functioning, DIF) (Camilli & Shepard, 1994). Claramente el DIF se presenta cuando no se satisface el supuesto de unidimensionalidad en tanto que existe otra variable (género, nacionalidad, nivel socio económico, la percepción del informante, etc.) ajena a la de interés que también contribuye en la explicación de la respuesta al ítem.

Lambert et al. (2007) compararon los resultados obtenidos con los adolescentes jamaquinos mediante la YSR con los datos pertenecientes a otro grupo de adolescentes alemanes. Tras estimar los parámetros de cada ítem para ambos grupos, ellos concluyeron que la invarianza de los parámetros entre estas dos culturas puede sostenerse aunque sólo parcialmente. Un 70% de los ítems de la dimensión Internalizante y un 65% del espectro Externalizante mostraron DIF según la nacionalidad ya sea porque diferían significativamente en los parámetros de localización o en los de discriminación. Sin embargo, si se cotejaban las Curvas Características de los Tests (funciones que surgen de combinar todas las CCI para cada dimensión) las diferencias se cancelaban, lo que apoyaría la invarianza de las mediciones a través de las culturas sólo si se utilizan todos los ítems.

Fuertemente vinculado a la noción de DIF aparece el concepto de sesgo. Mientras la primera es puramente estadística, el segundo considera las causas y busca explicaciones para tal funcionamiento diferencial. En su tesis doctoral, Ngo (2006) estudió el DIF de los ítems de seis escalas estrechas del CBCL en función de la raza. La autora concluyó que los ítems del CBCL así como algunas de sus escalas no pueden ser considerados equivalentes para registrar el comportamiento de niños de seis años afroamericanos y de sus pares euroamericanos. Para las escalas Ansioso/Depresivo, Problemas de Atención y Retraimiento, los padres de los niños afroamericanos tendieron a no percibir la presencia de síntomas (por lo que los valores de los parámetros de localización fueron más elevados que para los euroamericanos).

Ngo (2006) argumentó que son los factores contextuales los que explican este funcionamiento diferencial de los ítems y de las escalas. Las experiencias sociales y culturales, y las creencias de los padres sobre aquello que es o no un comportamiento normal, son variables que contribuyen a la percepción que tienen de la conducta de sus hijos. Por ende, las mismas constituyen una innegable fuente de sesgo. En definitiva, lo que muestra este estudio es que no es posible suponer una equivalencia conceptual de los constructos entre ambas culturas. Los problemas de comportamiento para los padres de una cultura no se relacionan con el constructo de la misma manera que en la otra cultura, lo cual demuestra que cada síntoma presenta un significado disímil en estas poblaciones (von Eye, 2004). Los resultados de Ngo (2006) abren un cuestionamiento respecto de la validez intercultural de las escalas del CBCL. Según esta investigadora, es más probable que la medición de las dimensiones del CBCL se vincule a factores del contexto cultural y se aleje de la presunción de universalidad.

### **CONSIDERACIONES FINALES**

El surgimiento y desarrollo formal de los modelos de la TRI se dio en el campo de la educación, y con el fin de medir el rendimiento. Aunque con cierta demora, en las últimas décadas se extendió su uso de manera progresiva hacia otras áreas y otros tipos de variables. Se espera que la exposición de las aplicaciones de la TRI a una prueba reconocida construida desde la teoría clásica contribuya a la difusión de su potencia y sus ventajas. Más aún en las aplicaciones a constructos propios del ámbito de la salud, donde la TCT es utilizada de forma casi exclusiva.

Para Pedreira et al. (1993) el avance en la investigación epidemiológica en salud mental infantojuvenil se conseguirá en la medida en que se profundice el estudio de las características psicométricas y la utilidad de escalas ya validadas como el CBCL y sus formas. En consonancia con este objetivo, la TRI ofrece un abordaje metodológico sofisticado que permite echar luz sobre aspectos de la validez interna-estructural del instrumento (Simms, 2008). La TRI no contradice ni desestima los resultados alcanzados desde el enfoque de TCT, sino que formula supuestos adicionales más exigentes para realizar un análisis más profundo y exhaustivo de la calidad y/o del funcionamiento del test (Hulin, Drasgow & Parsons, 1983).

Como se ha referido, la TRI además proporciona indicadores y procedimientos útiles para reducir la cantidad de ítems que componen las escalas del CBCL. De esta manera, se podría alcanzar una medición confiable en un menor tiempo de administración. Esto también ha sido destacado por Pedreira et al. (1993) como una necesidad propia de los instrumentos utilizados en la fase de cribado de las investigaciones epidemiológicas. Aho-

ra bien, a pesar de las herramientas técnicas que provee la TRI para acortar las escalas quizás sea conveniente resaltar que la reducción de los ítems no puede basarse exclusivamente en criterios estadísticos. La fortaleza de las escalas del CBCL está, justamente, en que permiten evaluar una amplia gama de psicopatologías. Por ende, el investigador deberá encontrar una solución adecuada para que la reducción no ponga en riesgo la validez de contenido del instrumento.

En las últimas décadas se ha registrado un creciente interés por el estudio transcultural de la morbilidad psiquiátrica tanto infantojuvenil como adulta. Pero también cabe destacar una preocupación creciente por optimizar los métodos utilizados para realizar comparaciones válidas y fiables. El análisis del DIF de los ítems del CBCL permite evaluar la calidad de las adaptaciones del formulario a otras poblaciones. La presencia del DIF dificulta la equiparación de las puntuaciones obtenidas por los miembros de cada grupo, lo cual es indispensable si se pretende cotejar los puntos de corte usados para determinar la prevalencia.

Asimismo, como se ha ejemplificado a partir de la investigación de Ngo (2006), el estudio del DIF permite evaluar la equivalencia conceptual del constructo a través de los grupos y generar hipótesis de interés psicológico referidas a la variable medida. Rojas y Pérez (2001) afirmaron que la TRI permitió una proliferación de métodos para la detección del DIF, sin embargo esto no contribuyó a la búsqueda de los factores explicativos del sesgo de los ítems. Smith (2002) recomienda a los investigadores revalorizar el concepto de sesgo por considerarlo una importante herramienta para la exploración del constructo. En este sentido, Lambert et al. (2007) afirmaron que estudios futuros con el CBCL deben ahondar en la comparación también de otros grupos; además de las diferencias transculturales, incluyeron las minorías étnicas, sociales y diferencias de género.

Otro aporte sustancial a los fines del modelo de Achenbach puede resultar del estudio del DIF según el informante que responda la prueba. La ausencia de correspondencia entre informantes en el ámbito de la evaluación de la psicopatología infantil es más la regla que la excepción (Godoy, 1996). Es justamente por esta razón que algunos autores defienden una evaluación multimétodo y multinformante (Achenbach & Rescorla, 2001; Valero, 1997). En cambio, otras posturas advierten que esta falta de concordancia entre las perspectivas de los informantes se sostiene en que se están evaluando dimensiones psicológicas diferentes que difícilmente puedan ser comparadas (Luciano, 1997). Frente a este panorama, muy probablemente el estudio del DIF según el informante abra un novedoso campo de discusión y nuevos desarrollos para mejorar este método de evaluación.

## REFERENCIAS BIBLIOGRÁFICAS

- Achenbach, T.M. & Edelbrock, C. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T.M. (1991a). *Manual for the Teacher's Report Form and 1991 profile*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T.M. (1991b). *Manual for the Youth Self-Report and 1991 profile*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T.M. & Rescorla, L.A. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington: University of Vermont, Research Center for Children Youth and Families.
- Camilli, G. & Shepard, L. (1994). *Methods for identifying biased test items*. Thousand Oaks: Sage.
- Cervone, N.; Samaniego, V.C.; Luzzi, A.M.; Slapak, S.; Padawer, M. et al. (2003). Enfoque epidemiológico en una población clínica de niños escolarizados. *Memorias de las Jornadas de Investigación en Psicología, Volumen X, 2*, 25 - 28.
- Curran, P.J.; Hussong, A.M.; Cai, L.; Huang, W.; Chassin, L.; Sher, K.J. et al. (2008). Pooling Data From Multiple Longitudinal Studies: The Role of Item Response Theory in Integrative Data Analysis. *Developmental Psychology, 44* (2), 365-380.
- Godoy, A. (1996). *Toma de decisiones y juicio clínico*. Madrid: Pirámide.
- Hulin, C.L.; Drasgow, F. & Parsons, C.K. (1983). *Item response theory: Application to psychological measurement*. Homewood: Dow Jones-Irwin.
- Ivanova, M.Y.; Achenbach, T.M.; Dumenci, L.; Rescorla, L.; Almqvist, F.; Bilenberg, N. et al. (2007). The Generalizability of the Youth Self-Report Syndromes Structure in 23 Societies. *Journal of Consulting and Clinical Psychology, 75* (5), 729-738.
- Kessler, R. & Mroczek, D. (1994). Scoring the UM-CIDI short forms. Ann Arbor: Universidad de Michigan, Institute for Social Research.
- Lambert, M.C.; Essau, C.A.; Schmitt, N. & Samms-Vaughan, M. E. (2007). Dimensionality and Psychometric Invariance of the Youth Self-Report Form of the Child Behavior Checklist in Cross-National Settings. *Assessment, 14*, 231-245.
- Lambert, M.C.; Schmitt, N.; Samms-Vaughan, M.E.; An, J.S.; Fairclough, M. & Nutter, C.A. (2003). Is it prudent to administer all items for each CBCL cross-informant syndrome? Evaluating the psychometric properties of the YSR dimensions via confirmatory factor analysis and item response theory. *Psychological Assessment, 15*, 550-568.
- Luciano, M.C. (1997). *Manual de Psicología Clínica. Infancia y adolescencia*. Valencia: Promolibro.
- Navarro, J.B.; Doménech, J.M.; de la Osa, N.; Ezpeleta, L. (1998). El análisis de curvas ROC en estudios epidemiológicos de psicopatología infantil: aplicación al cuestionario CBCL. *Anuario de Psicología, 29* (1), 3 - 15.
- Ngo, V.K. (2006). Contextual analysis of measurement bias in the Child Behavior Checklist for African American children. Tesis doctoral. Universidad de Vanderbilt.
- Pedreira, J.L.; Sanchez, B, Sardinero, L.; Martín, L. & Martín, P. (1993). Aplicación del método epidemiológico para la detección de trastornos mentales en la infancia y la adolescencia. En J.L. González de Rivera y Revuelta, F. Rodríguez Pulido & A. Sierra López, *El método epidemiológico en Salud Mental* (pp. 215-238). Barcelona: Ed. Masson-Salvat.
- Reiser, M. (1989). An application of the item response model to psychiatric epidemiology. *Sociological Methods & Research,*

18, 66-103.

- Rojas, A. & Pérez, C. (2001). *Nuevos Modelos para la Medición de Actitudes*. Valencia: Promolibro.
- Samaniego, V.C. (2008). El Child Behavior Checklist: su estandarización en población urbana argentina. *Revista de Psicología UCA*, 4 (8), 113-130.
- Samaniego, V.C. (2004). Prevalencia de Trastornos Psíquicos en población escolar de 6 a 11 años de edad. *Memorias de las XI Jornadas de Investigación en Psicología. Psicología, Sociedad y Cultura*. Buenos Aires, Tomo II, pp. 226-228.
- Simms, L.J. (2008). Classical and Modern Methods of Psychological Scale Construction. *Social and Personality Psychology Compass*, 2, 414 - 433.
- Smith, L.L. (2002). On the Usefulness of Item Bias Analysis to Personality Psychology. *Personality and Social Psychology*, 28, 754 - 763.
- Stansbury, J.P.; Ried, L.D. & Velozo, C.A. (2006). Unidimensionality and Bandwidth in the Center for Epidemiologic Studies Depression (CES-D) Scale. *Journal of Personality Assessment*, 86 (1), 10-22.
- Valero, L. (1997). La evaluación del comportamiento infantil: características y procedimientos. *Apuntes de Psicología*, 51, 51-79.
- von Eye, A. (2004). Commentaries: The Treasures of Pandora's Box. *Measurement*, 2 (4), 244-247.
- Zwinderman, A.H. (1990). The measurement of change of quality of life in clinical trails. *Statistics in Medicine*, 9, 931-942.

Fecha de recepción: 30 de marzo de 2010

Fecha de aceptación: 23 de agosto de 2010